Health Informatics Meets eHealth: Biomedical Meets eHealth – From Sensors to Decisions.

Günter Schreier
Dieter Hayn

HEALTH INFORMATICS MEETS eHEALTH

Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The series has been accepted by MEDLINE/PubMed, SciVerse Scopus, EMCare, Book Citation Index – Science and Thomson Reuters' Conference Proceedings Citation Index.

Volume 248

Recently published in this series

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

Health Informatics Meets eHealth

Biomedical Meets eHealth – From Sensors to Decisions

Proceedings of the 12th eHealth Conference

Edited by

Günter Schreier

AIT Austrian Institute of Technology GmbH, Graz, Austria

and

Dieter Hayn

AIT Austrian Institute of Technology GmbH, Graz, Austria

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-61499-857-0 (print)
ISBN 978-1-61499-858-7 (online)
Library of Congress Control Number: 2018941383

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

Preface

Biomedical Meets eHealth – From Sensors to Decisions

Already today, but even more in the future, ICT systems in healthcare and biomedical systems and devices will increasingly be intertwined and share a common entity, which is data. This completes the chain and flow of information from the sensor via the processing to the actuator, which can be anything from a human healthcare professional to a robot. Along this pathway, methods for automating the information processing, like signal processing, machine learning, predictive analytics and decision support, play an increasing role to provide actionable information and to support personalized and preventive health care concepts, in both biomedical and digital healthcare systems and applications.

Both scientific disciplines, i.e. biomedical engineering and health informatics, are also closely related to each other and it is often difficult to delineate where the one ends and the other begins.

This starts with the practical settings, for example, in hospitals. Traditionally, there are two different organizational entities working together, i.e. the healthcare engineering and the ICT departments. The first primarily take care of the "hardware", i.e. appliances, devices and systems, for example imaging equipment like Ultrasound machines or CT scanners. The latter take care of conveying the data generated by these items to the bedside and the healthcare specialists via Health Information Systems (HIS), Radiology Information Systems (RIS) or Picture Archiving and Communication Systems (PACS). However, these systems are interrelated and – as an example – it becomes more and more difficult to locate errors when combined systems come down. In the end, security is also a common concern and standards like the *IEC 80001: Application of risk management for IT-networks incorporating medical devices* address these overarching needs.

A look to the international level reveals, for example, that the IEEE Engineering in Medicine and Biology Society (EMBS),[1] the world's leading biomedical Engineering society, has named one of its flagship conferences "*Biomedical and Health Informatics*". As such, the name of this conference by itself very much integrates both worlds and in recent years this conference has been organized adjacent to the HIMSS Annual Conference on Healthcare IT, also evidence of the synergies that can be achieved by keeping these two topics close to each other. Therefore, it is fitting, that our opening keynote speaker, Prof. Nigel Lovell, has not only significantly contributed to both aspects in his professional and scientific life, but also currently happens to be the IEEE EMBS president.

Finally, since its beginning in 2007, the scientific backbone of our annual conference has been the working group for "Medical Informatics and eHealth" of the Austrian Society of Biomedical Engineering (OEGBMT)[2] and the Austrian Computer Society

[1] https://www.embs.org/
[2] http://www.oegbmt.at/arbeitsgruppen/medizinische-informatik-und-ehealth/

(OCG).[3] This unique group was founded by Prof. Günther Gell in 1981 and, since then, has been a forum for the many scientists, experts and practitioners with a stake in both worlds. As such, this year's subtitle was also motivated by making this long-lasting partnership more explicit.

Graz, March 2018

Günter Schreier
Dieter Hayn

[3] https://www.ocg.at/medizinische-informatik-und-ehealth

Scientific Program Committee

Univ.-Prof. Dr. Klaus-Peter Adlassnig, Medical University of Vienna, Austria
Prof. Dr. Josef Altmann, University of Applied Sciences Upper Austria, Wels, Austria
Univ.-Prof. Dr. Elske Ammenwerth, UMIT, Hall in Tirol, Austria
Univ.-Prof. Dr. Andrea Berghold, Medical University of Graz, Austria
Prof. Dr. Oliver Bott, University of Applied Sciences Hannover, Germany
Univ.-Prof. Dr. Ruth Breu, University of Innsbruck, Austria
Catherine Chronaki, HL7 Foundation, Brussels, Belgium
Univ.-Prof. DDr. Wolfgang Dorda, Medical University of Vienna, Austria
Prof. Dr. Stephan Dreiseitl, University of Applied Sciences Hagenberg, Austria
Univ.-Prof. Dr. Georg Duftschmid, Medical University of Vienna, Austria
Univ.-Prof. Dr. Martin Dugas, University of Münster, Germany
Univ.-Prof. Dr. Walter Gall, Medical University of Vienna, Austria
Dr. Holger Gothe, UMIT, Hall in Tyrol, Austria
Prof. Dr. Martin Haag, Heilbronn University, Germany
Prof. Dr. Peter Haas, University of Applied Sciences Dortmund, Germany
Prof. Dr. Anke Häber, University of Applied Sciences Zwickau, Germany
Dr. Dieter Hayn, MBA, AIT, Graz, Austria
Dr. Mira Hercigonja-Szekeres, University of Applied Sciences, Zagreb, Croatia
ao. Univ.-Prof. Dr. Alexander Hörbst, UMIT, Hall in Tirol, Austria
Prof. Dr. Ursula Hübner, University of Osnabrück, Germany
Sergey Karas, Siberian State Medical University Tomsk, Russia
Prof. Dr. Guido Kempter, University of Applied Sciences Vorarlberg, Austria
Prof. Dr. Peter Klutke, University of Kempten, Germany
Prof. Dr. Petra Knaup, University of Heidelberg, Germany
Prof. Dr. Werner Kurschl, University of Applied Sciences Hagenberg, Austria
Univ.-Prof. Dr. Richard Lenz, University of Erlangen-Nürnberg, Germany
Prof. Dr. Nicos Maglaveras, Aristotle University, Thessaloniki, Greece
Prof. Dr. Michael Marschollek, Medical University of Hannover, Germany
Prof. Dr. Christian Menard, University of Applied Sciences Carinthia, Austria
Prof. Dr. George Mihalas, University of Medicine and Pharmacy, Timisoara, Romania
Ass.-Prof. Dr. Ivana Ognjanovic, Donja Gorica University, Podgorica, Montenegro
Univ.-Prof. Dr. Karl-Peter Pfeiffer, University of Applied Sciences Joanneum, Graz, Austria
Univ.-Prof. Dr. Ulrich Prokosch, University of Erlangen, Germany
Dr. Stefan Sauermann, University of Applied Sciences Technikum Vienna, Austria
Prof. Dr. Paul Schmücker, University of Mannheim, Germany
Dr. Michael Shifrin, N.N. Burdenko Neurosurgical Institute, Moscow, Russia
Prof. Dr. Martin Stämmler, University of Applied Sciences Stralsund, Germany
Prof. Dr. Christian Stingl, University of Applied Sciences of Carinthia, Austria
Prof. Dr. Lăcrămioara Stoicu-Tivadar, Polytechnic University of Timisoara, Romania
Hanna Suominen, Adjunct Professor, Senior Researcher, PhD, MSc
Prof. Dr. Jana Zvarova, Charles University, 1st Faculty of Medicine, Prague, Czech Republic

Reviewers

We would like to thank all reviewers for their significant contribution to the proceedings of the eHealth2018:

Klaus-Peter Adlassnig
Jozef Aerts
Josef Altmann
Elske Ammenwerth
Mairanne Behrends
Andrea Berghold
Kerstin Denecke
Kerstin Denecke
Stephan Dreiseitl
Georg Duftschmid
Christo El Morr
Gerhard Fortwengel
Jan Gaebel
Walter Gall
Sebastian Garde
Holger Gothe
Martin Haag
Anke Häber
Emmanuel Helm
Mira Hercigonja-Szekeres
Michael Heß
Jürgen Holm
Ursula Hübner
Peter Kastner
Guido Kempter
Peter Klutke
Alexander Kollmann
Dagmar Krefting
Karl Kreiner
Johannes Kriegel
Martin Kropf
Werner Kurschl
Alexander Mense
Robert Mischak
Robert Modre-Osprian
Stefan Müller-Mielitz
Sabrina Neururer
Ivana Ognjanovic
Paul Panek
Hans-Ulrich Prokosch
Ida Pu
Christoph Rinner
Jean marie Rodrigues
Stefan Sabutsch
Stefan Sauermann
Hannes Schlieter
Paul Schmücker
Guenter Schreier
Eva Schulc
Stefan Schulz
Abbas Sheikhtaheri
Michael Shifrin
Stefan Skonetzki
Martin Staemmler
Christian Stingl
Lacramioara Stoicu-Tivadar
Sylvia Thun
Karsten Weber

Contents

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-1

The Effect of Latent Binary Variables on the Uncertainty of the Prediction of a Dichotomous Outcome Using Logistic Regression Based Propensity Score Matching

Szabolcs SZEKÉR[a] and Ágnes VATHY-FOGARASSY[a,1]
[a]*Department of Computer Science and Systems Technology, University of Pannonia, Hungary*

Abstract. Logistic regression based propensity score matching is a widely used method in case-control studies to select the individuals of the control group. This method creates a suitable control group if all factors affecting the output variable are known. However, if relevant latent variables exist as well, which are not taken into account during the calculations, the quality of the control group is uncertain. In this paper, we present a statistics-based research in which we try to determine the relationship between the accuracy of the logistic regression model and the uncertainty of the dependent variable of the control group defined by propensity score matching. Our analyses show that there is a linear correlation between the fit of the logistic regression model and the uncertainty of the output variable. In certain cases, a latent binary explanatory variable can result in a relative error of up to 70% in the prediction of the outcome variable. The observed phenomenon calls the attention of analysts to an important point, which must be taken into account when deducting conclusions.

Keywords. Case-Control Studies, Propensity Score, Logistic Regression, Uncertainty, Monte Carlo Method

1. Introduction

In clinical studies, cohort-based analyses are widely used to analyze the effect of specific risk factors, causes of diseases or complications of treatments. In these studies, characteristic data (e.g. patient history, applied treatments, medication, etc.) of two independent groups of patients are compared and the conclusions are drawn from the significant differences between the cohorts.

In cohort studies, the selection of cohorts is very important and has a significant impact on the output of the analysis as well. Individuals of these groups have to be similar in many ways (e.g. gender and age distribution, etc.), but they have to differ in a certain characteristic property (e.g. patients in the case group are treated with a certain medicine, while individuals in the control group receive placebo) [1]. In prospective study design,

[1] Corresponding Author: Ágnes Vathy-Fogarassy, Department of Computer Science and Systems Technology, University of Pannonia, 2. Egyetem Str., 8200 Veszprém, Hungary, E-Mail: vathy@dcs.uni-pannon.hu

there are many inclusion and exclusion criteria specified to select the proper individuals into the case and the control groups [2]. Patient-specific data, thought to be important and relevant, are systematically collected and recorded during the whole study period. The main disadvantage of these studies is that the execution of a study sometimes takes up a lot of time. In contrast, retrospective cohort studies look back in the time and they do not require a long time for collecting data about patients. However, these studies must face the fact, that the range of available data is not always complete. For this reason, it may happen that case and control groups differ not only in the previously planned characteristic property (e.g. medication treatment vs. placebo), but hidden differences may exist as well, for which we do not have data. Of course, similar cases may also occur in prospective studies, if the scope of data included in the study is not complete.

In the analysis phase, the effect of the known independent variables on the outcome can be determined in different ways. If the outcome variable (e.g. the appearance of a disease) is categorical, the logistic regression (LR) method is probably the most commonly used method for this purpose. Logistic regression determines the odds ratios for each explanatory variable and helps to explain relationships between the dependent variable (outcome) and the independent variables. However, the LR is based on estimation, and it includes uncertainty. The uncertainty of the model can be measured by R^2 measures, which suggest how well the observed outcome is replicated by the model from the independent variables. That is, if the established logistic regression model is inaccurate, then the value of the output variable (e.g. the appearance of a disease) cannot be predicted with sufficient certainty. However, the question arises, whether the uncertainty of the model can be derived from the latent variables. Furthermore, the uncertainty of the prediction of the output variable only arises from the predictive variables included in the model or the effect of the latent variables may also affect this uncertainty? How does the uncertainty of the model relate to the uncertainty of the prediction of the output variable?

In this paper, our goal is to present a statistics-based research in which we try to determine the relationship between the accuracy of the binary logistic regression model and the uncertainty of the prediction of the output variable. The analysis is based on benchmark datasets generated with Monte Carlo simulations. Using these datasets, binary logistic regressions based propensity score matching was performed under various conditions to generate possible control groups, and then the deviation of the output variable in the case and control groups was investigated in order to determine the degree of distortion. In the present study, only the effect of binary independent variables was analyzed.

The remaining part of this paper is organized as follows. Section 2 introduces the methods used during our study. In Section 3 the course of the research is presented, then Section 4 shows the results of the analysis. Finally, Section 5 concludes the paper.

2. Theoretical background

2.1. The methodology of the research

As we mentioned previously, the aim of our analysis is to measure the effect of the latent predictors to the outcome variable. The effect of the known predictive parameters on the dependent variable is expressed by the calculations of odds ratios using logistic regression. The inaccuracy of the prediction determined by the logistic model was

measured by the R^2 value of the model. This quantitative measure also estimates the extent of the deficit that comes from latent variables. After the creation of the binary logistic regression model, a propensity score value is calculated for each individual based on the odds ratios of the known predictive variables. Then, by the use of random sampling a case group is created and propensity score matching was performed to select the most proper individuals into a control group. Finally, the outcome variable of the control group and the case group were compared to measure bias between them and to estimate the effect of latent predictive variables.

These applied methods are introduced briefly in the next section.

2.2. Applied methods

2.2.1. Logistic Regression

Logistic Regression (LR) is a widely used regression model, ranging from machine learning to medical fields and social sciences, where the dependent variable is categorical. LR is an appropriate regression analysis to conduct when the dependent variable is dichotomous (binary): fit for healthcare-based studies where the outcome is the existence or lack of a diagnosis or condition. Using LR, it is possible to describe data and to explain relationships between the dependent variable (dichotomous outcome) and other independent variables, and aims to find the best fitting model to describe these aforementioned relationships by using a logistic function. LR estimates a multiple linear regression function defined as:

$$logit(p) = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k \tag{1}$$

where p is the probability of presence of the characteristic of interest and b_i is a regression coefficient indicating the relative effect of X_i explanatory variable on the outcome. The logit transformation is defined as the log odds:

$$odds = \frac{p}{1-p} = \frac{probability\ of\ presence\ of\ characteristics}{probability\ of\ absence\ of\ characteristics} \tag{2}$$

and

$$logit(p) = \ln\left(\frac{p}{1-p}\right) \tag{3}$$

The odds ratio for each independent variable is acquired by applying the exponential function to the corresponding factor.

2.2.2. Coefficient of determination

Logistic regression is a probabilistic model that does not guarantee that the regressed outcome is entirely describable with the independent variables: it is possible to measure this uncertainty and there are various methods to do so. The most basic measure is the coefficient of determination, denoted by R^2. R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variables. It provides a measure of how well a model approximates the observed outcomes based on the proportion of total variation of outcomes explained by the same model.

Usually, the value of R^2 is in the range of $[0, 1]$. The better the linear regression fits the data, the closer the value of R^2 is to 1. However, values of R^2 outside the range of $[0, 1]$ can occur, depending on the used measure.

R^2 does not indicate whether the independent variables cause the changes of the dependent variable or omitted-variable bias exists. There is no way to tell if the correct regression was used, if the most appropriate set of independent variables has been chosen or if there is a collinearity present in the data on the explanatory variables. The model might be improved by using transformed versions of the existing set of independent variables and it is possible that there are not enough data points to make a solid conclusion. It is important to take note of the second caveat: R^2 does not indicate whether omitted-variable bias exists. But still, R^2 provides a measure to quantify the model quality.

Our main research aim is to discover if there is a measurable numeric relationship or determined correlation between the value of general R^2 of models with omitted independent variables and the influence of the omitted independent variable on the outcome variable.

2.2.3. Propensity Score Matching

The Propensity Score (PS) is the probability of the outcome being true, based on the observed baseline characteristics. The propensity score is calculated for each individual independently. Patients with a near identical level of exposure will be assigned a similar propensity score. Matching based on the PS (Propensity Score Matching, PSM) allows one to reduce the effects of selection bias or confounding when estimating the effects of the outcome when using observational data [3, 4]. It is important to take into consideration the "no unmeasured confounders" assumption, stating, that all variables affecting treatment assignment and outcome have been measured [Rosenbam1983b]. There is no consensus in the literature as to which variables should be included in the propensity score model. There are merits in only including potential or true confounders [5]. Furthermore, there is no uniform agreement upon the definition of what constitutes as a maximal acceptable propensity distance (caliper size) between the matched individuals [6, 7]. Of course, if the caliper size is too large, then the accuracy of the matching-based selection decreases.

3. The applied methodology

Our study is based on the assumption that, if the set of observed variables is complete, the logistic regression model properly describes the relationship between the independent variables and the dependent variable and the R^2 value of the model is around 1.0. Selecting a control group to a sample based on this model guarantees that the deviation of the outcome variable is marginal between the case and the control groups based on the assumption that the odds ratio values are adequate. The effect of the latent variables was analyzed the following way. We omitted only one independent variable at a time from the dataset. On the resulting reduced dataset, we remodeled logistic regression, recalculated the propensity scores and evaluated the result using the methods described in the succeeding paragraphs. In short, we examined how omitted variables affect the R^2 value and how much deviation can be observed in the distribution of the output variable. The question is, therefore, that if we ignore an explanatory variable from the logistic

regression, then how can the R^2 value indicate the bias of the output variable (e.g. the incidence of a disease).

3.1. Simulation

The study was performed on benchmark datasets generated by Monte Carlo Simulation. Our investigation scenario consisted of 100 simulated datasets of size 1000 individuals characterized by 8 binary independent variables. All 8 independent variables $(x_1, \dots, x_8)$ were independent Bernoulli random variables with a probability parameter of 0.5. These independent variables model the characteristics of a certain patient, e.g. sex, diagnoses and other descriptors. For each dataset, additional datasets were created: for every independent variable, we omitted only one, while keeping the others intact. This resulted in 8 additional datasets, each one containing only 7 independent variables. This way the number of investigated datasets totals 900 $((1 + 8) * 100)$.

To determine the output variable we calculated a utility value (y') for each individuals based on Equation 4. As it can be seen, the values of the regression coefficients were chosen in such a way, that the effect of the independent variables changes uniformly. The binary outcome (coded by y) was determined by Equation 5 individually for each dataset. In a more comprehensible way, if the exposure of an element from a specific dataset was higher than the median of all elements from the same dataset, the outcome is 1 (having a diagnosis or receiving a treatment), otherwise 0. This way the probability of the outcome estimates 0.5 for each specific dataset.

$$y' = 1.0x_1 + 1.2x_2 + 1.4x_3 + 1.6x_4 + 1.8x_5 + 2.0x_6 + 2.5x_7 + 3.0x_8 \tag{4}$$

$$y = \begin{cases} 1 & if\ y' > \bar{y} \\ 0 & else \end{cases} \tag{5}$$

After the creation of the datasets, we performed logistic regression on each of them independently, to estimate the propensity scores of the individuals and the R^2 values for the models. In the next step, we determined the R^2 difference values for each coherent dataset by using Equation 6.

$$d_{R^2 i} = abs\left(R^2_{baseline} - R^2_{x_i}\right), i \in \{1, \dots, 8\} \tag{6}$$

where $R^2_{baseline}$ is the R^2 value of the dataset containing all independent variables and $R^2_{x_i}$ is the R^2 value of the dataset from which x_i was omitted.

In the next step, test groups were created by blind random selection from each dataset having the outcome retain 0.5 probability. The remaining elements formed the population, which contained the possible entities of the control groups. To create control groups, we ran propensity score matching (with caliper size=0.05) on each dataset 50 times and 50 possible control groups were selected for each test groups. The evaluation of the results was based on the average values of the 50 runs.

The individuals of the control groups were selected in two different ways. In the first case (*realistic case*), we assumed, that the population giving the basis of the control group contains individuals both with 1 and 0 values on the omitted binary variable. This case simulates when the population from which the control group is selected may contain random values on the invisible predictive parameter. In the second case (*pessimistic case*),

the worst case was modeled, when the population contains only such individuals where the value of the invisible predictive feature is equal to 1. This is the case, when we do not know, for example, that diabetes has a great impact on the outcome variable, and we select people into the control group without taking into consideration this feature, and the resulted control group contains only diabetic patients.

As the output variable in our study is binary, the distribution of the output was determined as the ratio of cases with $y = 1$ value, which models, for example, the frequency of a disease. So, the relative difference of the output variables in the case and control groups was calculated as follows:

$$rel\_err = \frac{frequency\ of\ elements\ in\ the\ control\ group\ with\ y{=}1\ outcome\ value}{frequency\ of\ elements\ in\ the\ case\ group\ with\ y{=}1\ outcome\ value} \quad (7)$$

Finally, we compared the calculated $d_{R^2 i}$ and rel_err values.

4. Results

During our work, we analyzed two possible scenarios. The first one is a realistic scenario, where the omitted variable in the population contained both 0 and 1 values, while the second one is a pessimistic scenario, where the omitted variable was uniform in the population with a value of 1.

Figure 1 shows the relationship between the probability of the outcome being 1 and the $d_{R^2 i}$ value. The left side of Figure 1 shows that there is no noticeable relationship between the accuracy of the logistic regression model and the probability of the outcome in the realistic scenario. The quality of the selected control groups is almost the same in every case. The deviation of the probability of the outcome from the expected value (shown in the figure as a cyan horizontal region which represents the minimum, average and maximum probability of the outcome being 1 calculated based on the case groups) is within a 10% range. It seems that the quality of the outcome variable is not affected

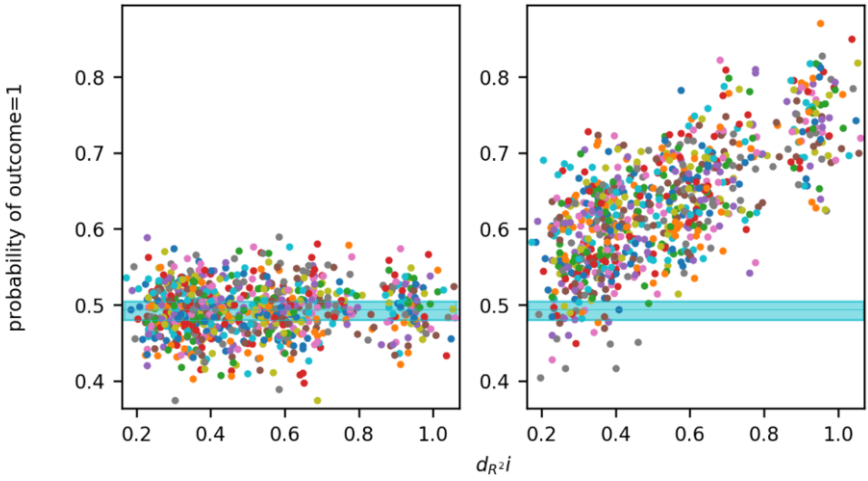


Figure 1. Relationship between the probability of the outcome being 1 and the $d_{R^2 i}$ values for the realistic scenario (left) and pessimistic scenario (right). There is a noticeable linear relationship on the right side, while this relationship is nonexistent on the left side.

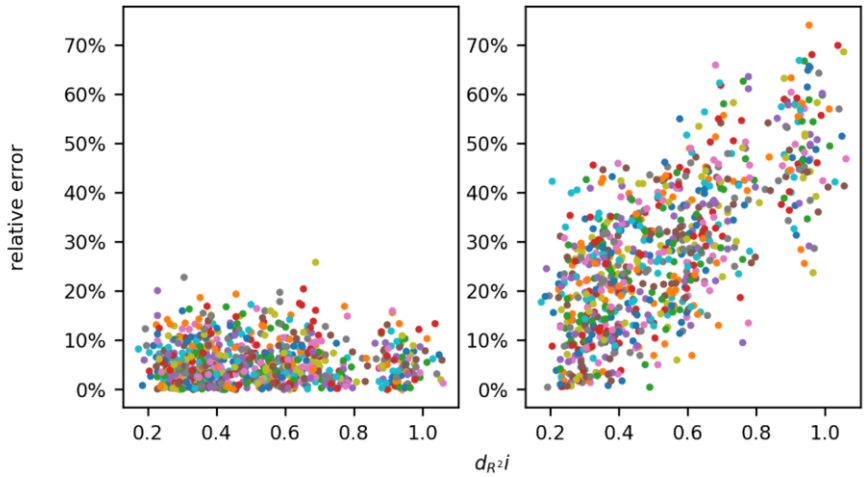


Figure 2. Relative error in the probability of the outcome being 1 as a function of d_{R^2i}. The same analogy applies in this case as well. On the left side (realistic scenario) there exists only a 20% relative error in the sample and the selected control, while on the right side (pessimistic scenario) this value rises up to 70% as a linear function of d_{R^2i}.

by the quality of the model. The right side of Figure 1 (pessimistic scenario) shows a more noticeable connection. The omitted variable strongly affects the value of the outcome variable. The worse the logistic regression model estimates the outcome, the bigger the difference is in the probability of the outcome variable. Namely, the probability of the outcome being 1 is a linear function of the inaccuracy of the logistic regression model. The more inaccurate the model, the higher the probability of the outcome being 1.

Figure 2. shows the relative error of the probability of the outcome being 1 between the case and the selected control groups as a function of $\mathrm{d_{R^2i}}$. Just as previously, on the left side (realistic scenario) there is no noticeable relationship and the relative error tops at 20%. In contrast, in the pessimistic scenario (right side) there is a linear relationship between the measures. The higher the inaccuracy of the model, the higher the relative error becomes, reaching even 70%.

Summarizing, the results of the logistic regression based analysis are significantly influenced by the ignoration of an explanatory binary variable. In the realistic case, when the omitted explanatory variable can take any value in the control group, the relative error of the predicted dichotomous value moves between 0% and 20%. However, if the omitted explanatory variable only takes 1 as value in the control group, the relative difference between the predicted dichotomous outcome value with an omitted explanatory variable and the outcome value without any omitted explanatory variable can reach 70%.

5. Discussion

The selection of independent variables is a critical step in case-control studies. The results of case-control studies rest on a correctly constructed dataset and adequate control group selection. In this paper, we analyzed the effect of the latent dichotomous variables

on the deviation of the outcome variable. The analysis was based on a Monte Carlo simulation in which we modeled the effect of omitted variables on the outcome. To measure the bias of the outcome we applied logistic regression based propensity score matching. We established that in the pessimistic scenario the omitted latent variables with high significance could greatly affect the value of the outcome variable. This conclusion is based on the revealed linear relationship between the deviance of the outcome and the model accuracy. This analysis draws attention to the important fact that calculations with latent variables can significantly influence the evaluation of case-control-based studies.

6. Acknowledgment

We acknowledge the financial support of ÚNKP-17-2 New National Excellence Program of the Ministry of Human Capacities and Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015.

7. References

[1] Everitt BS, Palmer CR: Encyclopaedic Companion to Medical Statistics, Hodder Arnold, London, 2005

[2] Sholom Wacholder, Joseph K. McLaughlin, Debra T. Silverman, and Jack S. Mandel: Selection of Controls in Case-Control Studies, Americal Journal of Epidemiology 135(9) (1992), 1019-1028.

[3] Austin PC: An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behavioral Research. 2011;46(3):399-424. doi:10.1080/00273171.2011.568786.

[4] Austin PC: Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples, Statistics in Medicine, 2011 May 20;30(11):1292-301. doi: 10.1002/sim.4200

[5] Austin, PC: Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: A Monte Carlo study. Statistics in Medicine, 26, 734–753. doi:10.1002/sim.2580

[6] Austin, PC: Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: A systematic review and suggestions for improvement. Journal of Thoracic and Cardiovascular Surgery, 134, 1128–1135. doi:10.1016/j.jtcvs.2007.07.021

[7] Austin, PC: The performance of different propensity score methods for estimating marginal odds ratios. Statistics in Medicine, 26, 3078–3094. doi:10.1002/sim.2781

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-9

A New Machine Learning Framework for Understanding the Link Between Cannabis Use and First-Episode Psychosis

Wajdi ALGHAMDI [a,1] , Daniel STAMATE [a], Daniel STAHL [b],
Alexander ZAMYATIN [c], Robin MURRAY [d] and Marta DI FORTI [e]
[a] *Data Science & Soft Computing Lab, and Department of Computing, Goldsmiths, University of London, UK*
[b] *Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK*
[c] *Faculty of Informatics, Department of Applied Informatics, National Research Tomsk State University, UK*
[d] *Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK*
[e] *MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK*

Abstract. Lately, several studies started to investigate the existence of links between cannabis use and psychotic disorders. This work proposes a refined Machine Learning framework for understanding the links between cannabis use and 1st episode psychosis. The novel framework concerns extracting predictive patterns from clinical data using optimised and post-processed models based on Gaussian Processes, Support Vector Machines, and Neural Networks algorithms. The cannabis use attributes' predictive power is investigated, and we demonstrate statistically and with ROC analysis that their presence in the dataset enhances the prediction performance of the models with respect to models built on data without these specific attributes.

Keywords. eHealth, Machine Learning, First-Episode Psychosis, Gaussian Processes, Support Vector Machine, Neural Networks.

1. Introduction

According to the World Health Organisation (WHO), eHealth is any secure, cheap, and efficient use of information and communications technologies in order to support health [1]. These days, more health care providers are replacing traditional paper notes with electronic patient records. In addition, the usage of advanced technologies such as computers, personal digital assistants, smart phones, etc. had enabled information to become more available and accurate. This lead to a tremendous increase in the electronic health data making an ideal promising land for applying machine learning algorithms to extract insights from data.

Currently machine learning algorithms are in the process of revolutionizing health. In just the same way as machine learning has made an enormous difference to business

[1] Corresponding author: Wajdi Alghamdi, Data Science & Soft Computing Lab, and Department of Computing, Goldsmiths, University of London, UK. Email: walgh001@gold.ac.uk

and industry, it will just as undoubtedly enhance medical research, and improve the practice of healthcare providers. For instance, machine learning algorithms have been successfully used in understanding the different manifestations of asthma [2], diagnosing psychosis [3][4], classifying leukaemia [5], detecting heart conditions in electrocardiogram (ECG) data [6], etc. In particular, machine learning algorithms have been proven to be capable of dealing with complicated medical data such as ECG signal data, where they show some outstanding results compared to traditional statistical approaches.

These studies suggest that machine learning can provide medical research with powerful techniques beyond the traditional statistical approaches mostly used, which include the conventional statistical tests, linear and logistic regression. Also, in biomedical engineering, several recent papers explored the potential for machine learning algorithms to detect various diseases. This has led to the publication of more guidance for medical researchers on how to infer and question such findings [7]. Finally, there is a tremendous interest in current interdisciplinary research into exploiting the power of machine learning to enable further progress in the new area of Precision Medicine, in which predictive modelling plays a vital role in forecasting treatment outcomes, and thus decisively contributes in optimising and personalising treatments for patients [8].

The medical field is considered as a critical area of research, yet there are still many difficult tasks that need to be carried out precisely and efficiently. The future success of health sector planning, and of healthcare in general, will be the adoption of intelligent systems where robotics and machine learning intersect. In order for health sector planning to catch up in this fast changing environment, machine learning must be put at the core of most strategies. For example, new developments in Psychiatry concern the so called Data-driven Computational Psychiatry, which relies heavily on the use of machine learning.

In this study, we propose a new computational psychiatry and machine learning framework based on developing optimised models for predicting the onset of first-episode psychosis with Gaussian Processes (GP), Support Vector Machines (SVM), and Neural Networks (NNET). In particular, our aim is the predictive modelling approach to help understanding the link between the first-episode psychosis and cannabis use. The dataset on which we based our study was collected by psychiatry practitioners and was used in previously conducted studies, such as [3][9]. It comprises an extensive set of variables, including demographic, drug-related and other variables, with specific information on participants' history of cannabis use - some of which being illustrated in Table 1.

The framework we present here integrates: data pre-processing, model tuning, model post-processing with receiver operating characteristic (ROC) optimisation based on the maximum accuracy cut-off threshold, and model evaluation with k-fold cross-testing. This sequence of enumerated phases is repeated 500 times for each GP, SVM with radial and polynomial kernels, and NNET algorithms, to study the potential variation of the performances of the resulting models. We investigated also the cannabis use attributes' predictive power by establishing statistically that their presence in the dataset augments the models' performances.

Table 1. Cannabis use attributes in the analysed dataset [3].

Attribute	Description
lifetime_cannabis_user	Ever used cannabis: yes or no
age_first_cannabis	Age upon first use of cannabis: 7 to 50
age_first_cannabis_under15	Age less than 15 when first used cannabis: yes, no or never used
age_first_cannabis_under14	Age less than 14 when first used cannabis: yes, no or never used
current_cannabis_user	Current cannabis user: yes or no
cannabis_fqcy	Pattern of cannabis use: never used, only on weekends or daily
cannabis_measure	Cannabis usage measure: none, hash less than once per week, hash on weekends, hash daily, skunk less than once per week, skunk on weekends, skunk daily
cannabis_type	Cannabis type: never used, hash or skunk
duration	Cannabis use duration: 0 to 41 (months)

2. Methods

2.1. Description of study population

The data we used to build our predictive models were a part of a case-control study [9]. The clinical data comprise 1106 records divided into 489 patients, 370 controls and 247 unlabelled records. The patients were individuals who presented with first-episode psychosis to the inpatient units of the South London & Maudsley Mental Health National Health Service (NHS) Foundation Trust. The controls were healthy people recruited from the same area served by the Trust. Each record in the data refers to a participant in the study and has 255 possible attributes divided into four groups. The first group consists of demographic attributes, which represent general features like gender, race and level of education. The second group of drug-related attributes contains information on the use of non-cannabis drugs, such as tobacco, stimulants and alcohol. The third group contains genetic attributes. These were removed from the analysis since they were out of this study scope. The final group contains cannabis-related attributes, such as the duration of use, initial date of use, frequency, cannabis type, etc.

For the purpose of building prediction models, we first removed any attribute with more than 50% missing value. Then, we perform the same high-level simplification of the dataset that was proposed in [3]. The resulting dataset, after the transformations, contained 783 records and 78 attributes. The records are divided into 451 patients and 332 controls. A summary of some of these fields—specifically, those that relate to cannabis use, such as type, age of first use and duration—can be seen in Table 1.

2.2. Data pre-processing

The quality of data may significantly affect the performance of the predictive models [10]. In order to help improve the quality of the data and, consequently, of the predictive models, the clinical data is pre-processed. Data pre-processing usually deals with the preparation and transformation of the initial dataset. In this study, we applied numerous pre-processing techniques such as missing values imputation, class balancing and feature selection to improve the efficiency and ease the modelling process.

Firstly, in term of missing values imputation, we applied random forest imputation from the *randomForest* package [11]. Although this method is computationally expensive, it enhanced the predictive power of the final models.

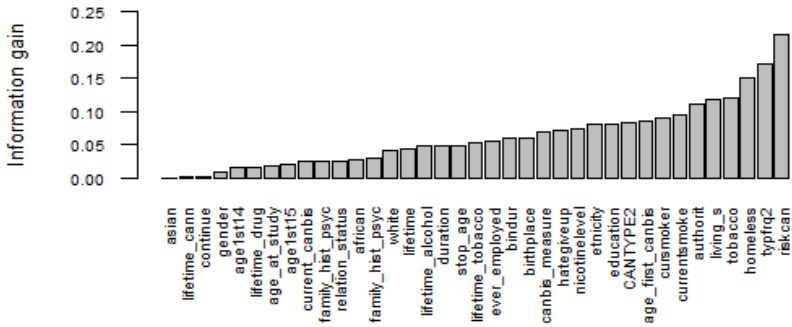


Figure 1. Attributes' predictive power with respect to Information Gain.

Secondly, the synthetic minority over-sampling technique (SMOTE) [12] was selected to treat the unbalanced classes existed in the data. SMOTE chooses a data point randomly from the minority class, de-terminus the K nearest neighbours to that point and then uses these neighbours to generate new synthetic data points using slight alterations. Our analysis used five neighbours. The results show that SMOTE leads to an increase in both the area under the ROC curve (AUC) and the accuracy.

Finally, we applied a feature selection technique based on the information gain [13]. To do so, we evaluate the information gain for each attribute with respect to the class. Such techniques are often used with forward selection or backward elimination, which considers only removing the feature subset with least ranking values. In this study, we apply information gain to filter out the attribute that does not have predictive power regarding information gain. Figure 1 illustrates some (due to lack of space) attribute predictive power with respect to the information gain. Figure 1 shows that attributes such as *riskcan* and *typefrq2,* which are cannabis measures are the highest ranked attribute when attributes such as gender are among the least ranked attributes. Initially, this indicates that some of the cannabis use attributes have more productivity power than some of the demographic attributes

2.3. Predictive modelling

To develop optimised predictive models for first-episode psychosis, we controlled the values of the parameters for each of the considered algorithms using chosen grids. Predictive models have been fitted in a five-fold cross-validation procedure, on each training set after pre-processing techniques were applied on the same training set, and have been tested on each test set. Models based on SVM, NNET, and GP were optimised to maximise AUC.

First, SVM models were tuned with different kernels such as SVM with the radial kernel (SVMR) and SVM with the polynomial kernel (SVMP). The optimal SVM models were obtained with SVMR, after tuning the parameters cost and gamma over 10 values. The optimal values for cost and gamma were 16 and 0.004, respectively.

Then, GP models were tuned with different kernels such as GP with the radial kernel (GPR) and GP with the polynomial kernel (GPP). The optimal GP models were obtained with GPP, with the parameters degree and scale, tuned over 10 values. The optimal values for degree and cost were 3 and 0.01, respectively.

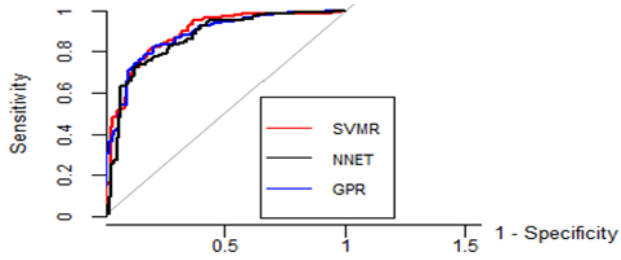


Figure 2. ROC curves for 3 models: SVMR, NNET and GPR.

Finally, NNET models were tuned over 15 values for the size (i.e. the number of hidden units) and 15 values for the decay (i.e. the weight decay), which is the parameter in the penalisation method for model regularisation. The optimal values were 13 and 0.01, respectively.

2.4. Predictive model post-processing

ROC curves allow visual analyses of the trade-offs between a predictive model's sensitivity and specificity regarding various probability cut-offs. The curve is obtained by measuring the sensitivity and specificity of the predictive model at every cutting point and plotting the sensitivity against 1-specificity. Figure 2 shows the ROC curves obtained for three of our predictive models, which are SVMR, NNET and GPR. The curve shows that SVMR performs better than other models regarding the evaluation dataset.

Several methods exist for finding a new cut-off threshold on the ROC curve. In this study, we find the point on the ROC curve corresponding to the highest accuracy.

2.5. Overall modelling procedure

The overall modelling procedure, which is based on data pre-processing, model optimisation, model post-processing and model evaluation, is inspired by [4] and reformulated and adapted to the context of the present framework. First, the dataset is randomly split, with stratification, in 60% and 40% parts denoted here by P1 and P2, respectively. Then, P1 is used for training and for optimising the model, as explained in Subsection 2.3. Different pre-processing methods that were explained in Subsection 2.2 were appropriately integrated into the cross-validation. Finally, in order to further enhance the model performance, the post-processing and model evaluation methods were applied to the optimised model using k-fold cross testing on the P2 dataset. In the k-fold cross testing procedure, we produce k post-processed model variants of the original optimised model. First, we create k stratified folds of the P2 dataset. Then, k-1 folds are used to find an alternative probability cut-off, that corresponds to the highest accuracy, on the ROC curve. The remaining one-fold is scored with the post-processed model based on the newly found cut-off point. This procedure enhances the predictive models and ensures that the datasets for post-processing and scoring are always distinct.

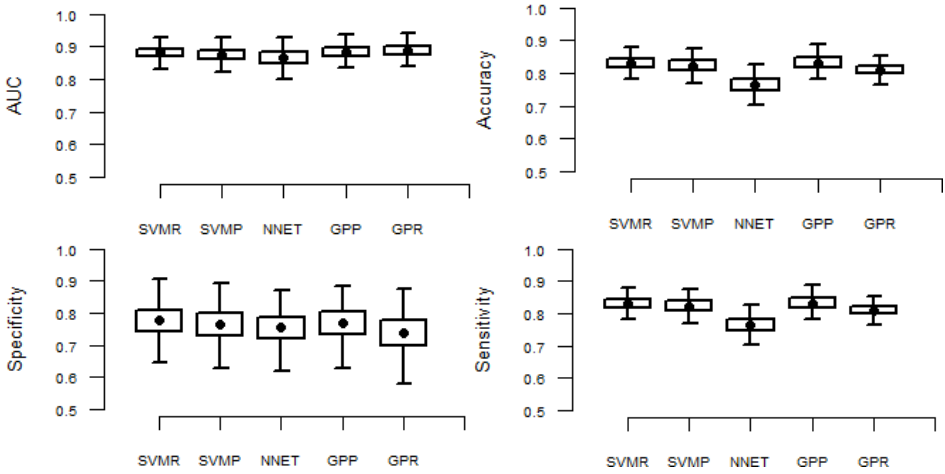


Figure 3. 500 repeated experiments simulations on Support Vector Machines with Radial (SVMR) and Polynomial kernels (SVMP), Gaussian Processes with Radial (GPR) and Polynomial kernels (GPR) and Single Layer Neural Networks (NNET).

3. Results

Due to expected potential variations of the predictive models' performance, we conducted extensive repeated experiments simulations to study these variations and the models' stability. The simulations consisted of 500 iterations of the procedure explained in Subsection 2.5. The models' performances concerning accuracy, AUC, sensitivity and specificity were evaluated for each iteration.

The aggregation of all iterations yielded various distributions of the above performance measures. These distributions were then visualised using box plots in Figure 3 to capture the models' performance capability and stability.

Also, estimations of the predictive neural networks' performances regarding means and standard deviation (SD) are shown in Table 2. We report results regarding models which are post-processed with ROC optimisation based on the largest accuracy cut-off method, as explained in Subsection 2.5. The results show that SVMR achieved the best results with a mean accuracy of 0.83 (95% CI [0.79, 0.87]) and a mean sensitivity of 0.87 (95% CI [0.81, 0.93]), similar to the results achieved by GPP. The rest of the predictive models scored a mean accuracy of 81%, which is better than all performances scored at [3].

Overall, we find that the models, especially SVMR and GPP, have good predictive power and stability, based on an acceptable level of variation in their performance measures evaluated across extensive repeated experiments simulations. Also, the results indicate that the performance differences between the different methods for selecting the ROC cutting points are not significant regarding the 4 performances.

After performing the repeated experiments simulations, we further investigated the predictive models in order to better comprehend the predictive power of the cannabis-related attributes over first-episode psychosis via statistical tests. To do so, we re-fit our performing models but removed the cannabis-related attributes, represented in Table 1, from the dataset. Then, we compared the performances of the models built with and without the cannabis-related attributes using t-test. We thereby demonstrated the

Table 2. Estimations of the predictive model's performances.

Model	Accuracy		AUC		Sensitivity		Specificity	
	Mean	SD	mean	SD	mean	SD	mean	SD
SVMR	0.83	0.02	0.88	0.02	0.87	0.03	0.77	0.04
SVMP	0.81	0.02	0.87	0.02	0.81	0.04	0.80	0.04
NNET	0.81	0.02	0.86	0.03	0.81	0.04	0.80	0.04
GPR	0.81	0.02	0.86	0.04	0.86	0.03	0.73	0.03
GPP	0.83	0.02	0.89	0.02	0.88	0.03	0.77	0.04

predictive value of cannabis-related attributes with respect to first-episode psychosis by showing that there is a statistically significant difference between the performances of the predictive models built with and without the cannabis variables.

Our analysis showed that the accuracy of SVMR decreased by 6% if the cannabis-related attributes were dropped from the process of building the predictive models as shown in the right image in Figure 4. If we compare, for instance, the accuracies of the SVMR models built on the data sets with and without the cannabis use attributes, the p-value obtained for the one-tailed t-test was 0.0002. This means that the predictive models with cannabis attributes have higher predictive accuracy than the models that were built without the cannabis attributes. This leads us to conclude that the additional cannabis variables jointly account for predictive information over first-episode psychosis. These results are consistent with findings from [3]. Also, we demonstrated that there is a significant difference between the ROC curves of the predictive models built with and without the cannabis variables as shown in the lift image in Figure 4. This also confirms the idea that the predictive models with cannabis attributes have higher predictive power than the models that were built without the cannabis attributes.

4. Conclusion

The advent of machine learning has so far proved to be of prime importance and capability in various fields, and recently in medical research and healthcare. This paper proposes a novel computational psychiatry and machine learning framework based on developping predictive models for the onset of first-episode psychosis in presence of clinical data including also cannabis use. We explored three types of machine learning

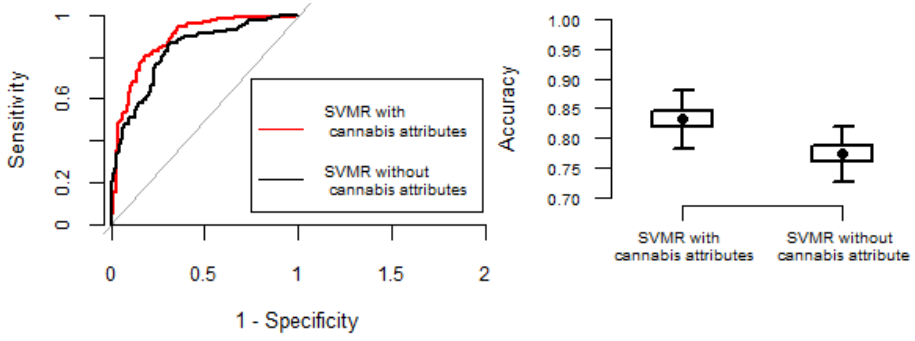


Figure 4. Left: ROC curves for optimised SVMR, with and without the cannabis attributes. Right: boxplots for 500 repeated experiments simulations for optimised SVMR, with and without the cannabis attributes.

algorithms, namely Gaussian Processes, Support Vector Machines, and feed-forward neural networks. Models are tuned and further optimised via post-processing, and evaluated with a k-fold cross testing methodology. In order to study the variation of the performances of the prediction models, the framework incorporates 500 repetitions of the model building, optimising, testing sequence. Experimental results show that the 3 machine learning algorithms lead to comparable models, with a slight advantage for Support Vector Machines and Gaussian Processes in front of neural networks.

Our best models score an average accuracy of 83%, which is above all the accuracy performances achieved in previous studies such as [3]. This paper extends on previous work as [3] by proposing a new machine learning framework based on a novel methodology in which models are post-processed based on ROC optimisation, and evaluated with the recent method of k-fold cross testing which we adapt after [4]. Moreover, in this new methodology, we developed optimized models with other powerful techniques such as Gaussian Processes and artificial neural networks not addressed in [3]. We also demonstrate statistically that the best models' performance decreases if cannabis attributes are removed from the analysis. This fact is also confirmed and illustrated by ROC analysis.

References

[1] World Health Organization. [online] Available at: http://www.who.int/en/ [Accessed 10 Feb. 2018].
[2] D. Belgrave, R. Cassidy, D. Stamate, et al., Predictive Modelling Strategies to Understand Heterogeneous Manifestations of Asthma in Early Life, 16th IEEE International Conference on Machine Learning and Applications, 2017.
[3] W. Alghamdi, D. Stamate, et al., A Prediction Modelling and Pattern Detection Approach for the First Episode Psychosis Associated to Cannabis Use, 15th IEEE International Conference on Machine Learning and Applications, 2016. pp.825-830.
[4] A. Katrinecz, D. Stamate, et al., Predicting Psychosis Using the Experience Sampling Method with Mobile Apps, 16th IEEE International Conference on Machine Learning and Applications, 2017.
[5] M. Adjouadi, M. Ayala, et al., Classification of Leukaemia Blood Samples Using Neural Networks. Ann Biomed Engineering, 2010. 1473-82.
[6] Y. Yan, X. Qin, et al., A Restricted Boltzmann Machine Based Two-Lead Electrocardiography Classification, in Proc. 12th Int. Conf. Wearable Implantable Body Sens, 2015.
[7] K. Foster, R. Koprowski, et al., Machine learning, medical diagnosis, and biomedical engineering research-commentary, BioMedical Engineering OnLine, 2014. pp. 94-95.
[8] R. Iniesta, D. Stahl, and P. McGuffin, Machine learning, statistical learning and the future of biological research in psychiatry, Psychological Medicine, 2016. pp. 2455–2465.
[9] M. Di Forti, A. Marconi, et al., Proportion of Patients in South London with First-Episode Psychosis Attributable to Use of High Potency Cannabis: a Case-Control Study, The Lancet Psychiatry, 2015.
[10] M. Kuhn, K. Johnson, Applied Predictive Modelling. Springer, 2013.
[11] A. Liaw, M. Wiener, Classification and Regression by randomForest, R News 2(3), 2002.
[12] N. Qazi, K. Raza, Effect of Feature Selection, SMOTE and Under Sampling on Class Imbalance Classification, 2012 UKSim 14th, 2012. pp. 145-150.
[13] P. Tan, M.Steinbach, V.Kumar, Introduction to Data Mining, 2016.
[14] M. Pepe, The Statistical Evaluation of Medical Tests for Classification and Prediction, New York: Oxford University Press, 2003.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-17

Separating Business Logic from Medical Knowledge in Digital Clinical Workflows Using Business Process Model and Notation and Arden Syntax

Jeroen S. DE BRUIN[a,b], Klaus-Peter ADLASSNIG[b,c,1], Harald LEITICH[d], and Andrea RAPPELSBERGER[c]

[a] *Institute eHealth, Department of Applied Computer Sciences, FH Joanneum University of Applied Sciences GmbH, Graz, Austria*
[b] *Medexter Healthcare GmbH, Vienna, Austria*
[c] *Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria*
[d] *Department of Obstetrics and Gynecology, Vienna General Hospital and Medical University of Vienna, Austria*

Abstract. Background: Evidence-based clinical guidelines have a major positive effect on the physician's decision-making process. Computer-executable clinical guidelines allow for automated guideline marshalling during a clinical diagnostic process, thus improving the decision-making process. Objectives: Implementation of a digital clinical guideline for the prevention of mother-to-child transmission of hepatitis B as a computerized workflow, thereby separating business logic from medical knowledge and decision-making. Methods: We used the Business Process Model and Notation language system Activiti for business logic and workflow modeling. Medical decision-making was performed by an Arden-Syntax-based medical rule engine, which is part of the ARDENSUITE software. Results: We succeeded in creating an electronic clinical workflow for the prevention of mother-to-child transmission of hepatitis B, where institution-specific medical decision-making processes could be adapted without modifying the workflow business logic. Conclusion: Separation of business logic and medical decision-making results in more easily reusable electronic clinical workflows.

Keywords. Guideline Adherence; Decision Support Systems, Clinical; Automatic Data Processing; Hepatitis B; Obstetrics and Gynecology Department, Hospital

1. Introduction

Clinical guidelines contain detailed instructions for the diagnosis and treatment of specific diseases or coping with difficult clinical situations. Their aim is to guide physicians in the decision-making process and introduce standard-based patient care [1]. In the ideal case, clinical guidelines are evidence-based, thereby identifying and integrating the most recent advances in prevention, diagnosis, treatment, prognosis, and cost-effectiveness in patient care [2]. As such, clinical guidelines have great potential to improve quality management and patient safety [3].

[1] Corresponding Author: Univ.-Prof. DI Dr. Klaus-Peter Adlassnig, Medexter Healthcare GmbH, Borschkegasse 7/5, A-1090 Vienna, Austria, E-Mail: kpa@medexter.com

Although evidence-based clinical guidelines are accepted among patient caregivers, they are frequently neglected in clinical routine because of their poor accessibility or time constraints on the part of clinicians. One solution to these limiting factors is to integrate clinical guidelines as computerized workflows in hospital or departmental information systems [1].

Clinical guidelines are usually documented as narratives containing descriptive texts, partially structured if-then statements, annotated decision tables, and/or narrated decision trees. In order to be integrated into an information system and processed automatically, the narratives need to be transformed into a standardized representation language and compiled into a computer-interpretable program. Several guideline models and execution engines are available for clinical use [4, 5].

One property of most guideline models and execution engines is that medical knowledge is combined with business logic. This restricts their reusability because healthcare institutions may have the same or similar workflows, but employ different underlying medial decision-making processes. We decoupled a clinical workflow's business logic, i.e. logic describing *which actions to undertake and when* during a diagnostic process, from medical decision-making, i.e. medical knowledge on *how to perform these actions*.

We adopted the Business Process Model and Notation (BPMN), version 2.0, Object Management Group (OMG) standard [6] to represent clinical guidelines in abstracted workflow form, and the Health Level Seven (HL7) knowledge representation standard Arden Syntax [7] to represent digitized medical knowledge. BPMN is a graphical representation language designed to model business processes and transform them into a set of activities. Several authors who reported on the implementation of clinical guidelines in BPMN noted significant improvements in their decision-making processes [8-10]. Arden Syntax is an international standard for medical knowledge representation and processing, extensively applied for clinical decision support (CDS).

The aim of our project is to devise a set of best-practice methods on the implementation and deployment of digital clinical guidelines in order to ensure and optimize the quality of patient care. In the present study, we focus on an important subproblem, namely how business logic can be separated best from medical decision-making when constructing an electronic clinical guideline. As a use case, we implemented a clinical workflow from the department of obstetrics and gynecology at the Vienna General Hospital (VGH), which contains evidence-based instructions on how to prevent mother-to-child transmission of hepatitis B.

2. Methods

This study addresses the implementation of the "Check for hepatitis B status at an outpatient visit of pregnant women" (HepBPW) guideline. This clinical guideline provides evidence-based instructions on how to prevent mother-to-child transmission of hepatitis B during pregnancy, and is part of the quality management framework at the department of obstetrics and gynecology, Vienna General Hospital [11].

The HepBPW guideline describes how outpatients are tested for their hepatitis B status, using hepatitis B antigen test results as well as quantitative hepatitis B polymerase chain reaction (PCR) test results. Based on the availability and outcome of these results, patients are instructed about breastfeeding (Table 1) and coached on adhering to these recommendations (such as how to wean a newborn from breastfeeding).

Table 1**.** Hepatitis B breastfeeding recommendations for HBsAg-positive patients. Table adapted from [11].

Test results		Hepatitis B PCR		
		Positive[1]	**Negative**[2]	**Unknown**
HBeAg	**Positive**	Don't breastfeed or wean from breastfeeding	Breastfeeding after immunization possible	Don't breastfeed or wean from breastfeeding
	Negative	Don't breastfeed or wean from breastfeeding	Breastfeeding after immunization possible	Breastfeeding after immunization possible
	Unknown	Don't breastfeed or wean from breastfeeding	Breastfeeding after immunization possible	No recommendation possible until data is available

[1] *Positive corresponds to* $>2*10^6$ *genomes/ml*
[2] *Negative corresponds to* $\leq 2*10^6$ *genomes/ml*
Note: HBeAg, hepatitis B envelope antigen; PCR, polymerase chain reaction

2.1. Data Collection and Storage

For workflow development and testing purposes, patient cases covering all possible guideline scenarios were provided by the department of obstetrics and gynecology at the VGH; these data were stored in a MySQL database.

2.2. Business Process Model and Notation

To design BPMN-based workflows we used core elements of the BPMN standard, which include *flow objects*, *connecting objects*, *swim lanes*, and *artefacts*.

Flow objects are the main elements describing workflow, and consist of *events*, *activities*, and *gateways*. Events are basic elements that trigger or terminate a process, such as start and stop events. Activities constitute an actual sequence of tasks that need to be executed in the process. Finally, gateways are constructs that determine control flow.

Flow objects are joined by connecting objects. They comprise *sequences*, *messages*, and *associations*. A sequence flow shows the order in which activities are performed, a message flow indicates the messages exchanged between workflow participants, and associations associate an artefact or text with a flow object.

Two swim lane components exist: *pools* and *lanes*. These serve as visual mechanisms that organize and categorize activities. A pool consists of major participants in a process, such as different organizations or different organizational branches. Lanes are used to organize and categorize activities within a pool.

Artefacts enable developers to annotate models with additional information in order to improve their readability. These comprise *data objects*, *groups*, and *text annotations*. Data objects show which data are needed or produced in an activity. Groups are used to categorize different activities without affecting the model's flow. Finally, text annotations are used to improve the readability and comprehensibility of the model.

We employed the open-source Activiti BPMN 2.0 Platform for the development of BPMN workflows, This framework offers the following components among others [12, 13]:

- the Activiti Engine, a Java process engine that runs BPMN 2.0-defined workflow processes;
- the Activiti Explorer, a web application that provides access to the Activiti Engine runtime;

- the Activiti Designer, an Eclipse plugin which permits the operator to graphically model BPMN 2.0 compliant workflow processes.

2.3. Arden Syntax and ARDENSUITE

We used Arden Syntax for the implementation of medical data access and knowledge-based clinical rule evaluation. Arden Syntax is a knowledge representation and processing standard capable of computerized representation and processing of medical knowledge, such as rules, decision trees, tables, and scores. Constituent medical rule sets and procedures are known as medical logic modules (MLMs) [14], and usually contain sufficient logic to make at least a single medical decision. Such a decision may result in a clinical alert, reminder, recommendation, or a circumscribed activity in a clinical guideline [15-17].

We used the ArdenSuite CDS technology platform for the execution of Arden Syntax MLMs, [18]. ARDENSUITE consists of an ARDENSUITE server with an Arden Syntax engine and a connector to relational databases, which is used for the storage, management, and execution of MLMs. Furthermore, it contains an ARDENSUITE integrated development and test environment (IDE), which serves as an authoring and test tool for Arden Syntax MLMs. Bidirectional access to MLMs is provided via representational state transfer (REST) or simple object access protocol (SOAP) requests and responses [18, 19].

2.4. Electronic Workflow Construction Principles and Restrictions

A guiding principle in constructing electronic workflows was to retain the original structure of the workflow and not introduce optimizations. In doing so, the workflow remained recognizable to our clinical partners, which promotes the acceptance of our method among clinicians. Furthermore, the workflows are implemented in such a way that they are not directly dependent on patient data, but rather on classification labels, thus improving their readability and reusability. The following guidelines and restrictions were used in modeling:

1. Every medical *action*, such as test ordering, test result communication, or medical examination, is by default modeled as a *user action*.
2. Every non-medical process, such as data preparation, abstraction etc., is implemented as a *service task.*
3. Every medical *decision*, such as decisions and labels generated on patient data to advance patient diagnosis or treatment, is implemented in Arden Syntax and implemented as a *remote service task*.

3. Results

Figure 1 shows the Activiti implementation of the HepBPW guideline.

The contents of the workflow results include instructions for:

- Active and passive immunization after childbirth, if necessary
- Specific breastfeeding recommendation according to Table 1
- Referrals for hepatitis B antigen and/or hepatitis B PCR testing

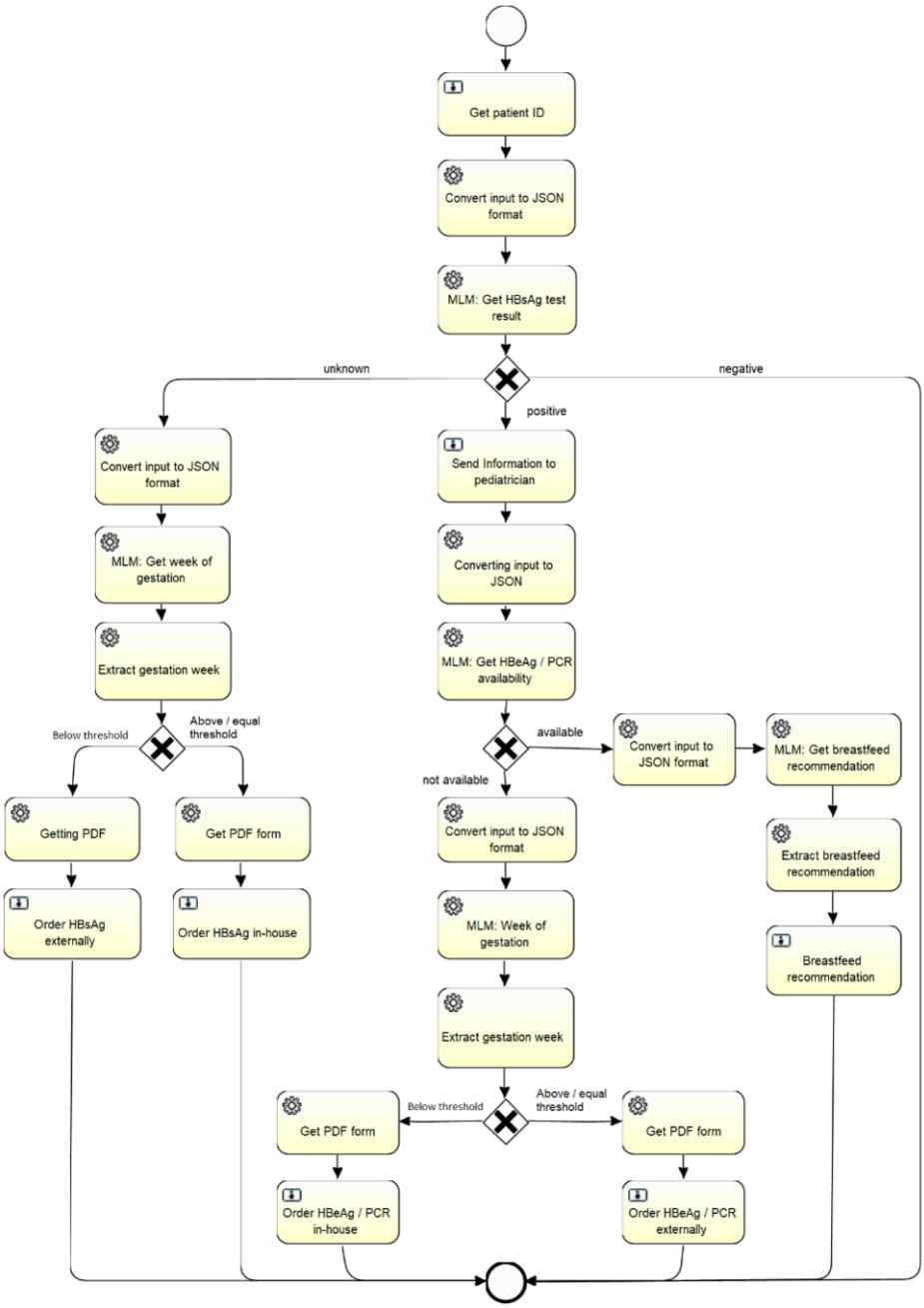


Figure 1. Activiti BPMN workflow for the HepBPW clinical workflow.

For the clinical use case, we constructed four MLMs that were responsible for clinical data retrieval and interpretation, including the hepatitis B breastfeeding recommendations shown in Table 1. Table 3 provides an overview of these MLMs and their descriptions.

Table 2. Activiti service tasks created for the clinical use case and their description.

Activiti service task	Description
Convert input to JSON format	Converts a string input to JSON format
Extract gestational week	Extracts a patient's gestational week interpretation
Extract breastfeeding recommendation	Extracts breastfeeding recommendations for a patient
Get PDF form	Retrieves an institute-specific PDF form
Tasks with prefix *MLM:*	Calls institute-specific evaluations implemented in Arden Syntax. These activities function as a wrapper for a REST call to the ARDENSUITE

Finally, in order to orchestrate uniform interaction between Activiti and the ARDENSUITE, we developed a service in Java that can call MLMs over REST communication and receive the results of an MLM call. In this service, parameters and responses are both formatted in JavaScript object notation (JSON).

Table 3. MLMs for the provision and interpretation of medical data related to the HepBPW workflow.

MLM name	Description
BreastfeedDecision	Implements the hepatitis B breastfeeding recommendation as stated in Table 1. The parameters for this decision are read from institute-specific medical data repositories.
GestationWeek	Reads the patient's gestational week from the medical database and determines whether it is below or above an institute-specific threshold related to an action in the workflow.
HBeAg-PCR-Availability	Checks in the institute-specific medical data repositories whether hepatitis B antigen and/or hepatitis B PCR test results are available for a patient.
HBsAgResult	Reads the patient's hepatitis B surface antigen test result from the institute-specific medical data repositories.

Note: HBsAg, hepatitis B surface antigen; HBeAg, hepatitis B envelope antigen; PCR, polymerase chain reaction

4. Discussion

We presented a clinical workflow solution that combines a BPMN-based workflow model with Arden Syntax MLMs. This solution permits the simplification of complex clinical guidelines by dividing them into an evidence-based general workflow layer, and an institute-specific medical decision-making layer. It is an essential step for integrating evidence-based clinical guidelines into the patient care process, and aids in the widespread dissemination and use of such guidelines.

Compared to other reports on the use of BPMN for modeling clinical pathways [8-10], we present a more transparent model of individual clinical workflow parts. Moreover, our method facilitates CDS.

As we developed our solution in a controlled environment, several limitations need to be considered. First, the solution needs to be evaluated by different stakeholders because the guideline was not tested across different clinical departments. As this is only a preliminary technical and clinical feasibility study, an acceptance study is yet to be performed. Finally, as the workflow contains sensitive patient data, the safety and security of data communication between platforms need to be assessed.

As clinical guidelines tend to be neglected in clinical routine, this approach is an innovative milestone in the electronic marshalling of such guidelines with the purpose of achieving the highest quality of patient care. The next step will be to test this approach

in a real-time clinical environment and study workflow results directly in clinical routine; previous projects have already shown how Arden Syntax MLMs can be directly integrated into clinical routine [20, 21].

5. Conclusion

Creating a BMPN-based clinical workflow in combination with an Arden-Syntax-based rule engine proved to be an efficient method of representing and automatically processing clinical guidelines. The use of Activiti enabled us to implement the content of an evidence-based clinical guideline as a BPMN-based clinical workflow, while the ARDENSUITE server and rule engine software instantiates the necessary data access and knowledge-based tasks. This solution permits the implementation of human-readable guidelines and sets a milestone for further clinical informatics research as well as the implementation of clinical guidelines and processing using established standards from two organizations, namely OMG and HL7 International.

References

[1] Seitinger, A, Rappelsberger, H. Leitich, M. Binder, K.-P. Adlassnig, Executable medical guidelines with Arden Syntax—Applications in dermatology and obstetrics, Artif Intell Med (2018), doi: 10.1016/j.artmed.2016.08.003.
[2] Masic, M. Miokovic, B. Muhamedagic, Evidence based medicine – New approaches and challenges, Acta Inform Med 16 (2008), 219–225.
[3] D.F. Lobach, W.E. Hammond, Computerized decision support based on a clinical practice guideline improves compliance with care standards, Am J Med 102 (2008), 89–98.
[4] M. Peleg, Computer-interpretable clinical guidelines: A methodological review, J Biomed Inform 46 (2013), 744–763.
[5] D. Isern, A. Moreno, Computer-based execution of clinical guidelines: A review, Int J Med Inform 77 (2008), 787–808.
[6] Business Process Model and Notation (BPMN), http://www.omg.org/spec/BPMN/2.0/PDF/, last access: 1.1.2018.
[7] P. Haug, R.A. Jenders, K.-P. Adlassnig, K. Fehre, S. Kim, C.Y. Jung, Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10, November 2014, Health Level Seven International, http://www.hl7.org/implement/standards/product_brief.cfm?product_id=372, last access: 1.1.2018.
[8] S. Rodriguez-Loya, A. Aziz, C. Chatwin, A service oriented approach for guidelines-based clinical decision support using BPMN, Stud Health Technol Inform 205 (2014) 43–47.
[9] M.G. Rojo, E. Rolón, L. Calahorra, F.Ó. García, R.P. Sánchez, F. Ruiz, N. Ballester, M. Armenteros, T. Rodriguez, R.M. Espartero, Implementation of the Business Process Modelling Notation (BPMN) in the modelling of anatomic pathology processes, Diagn Pathol 3(Suppl 1) (2008), S22.
[10] E. Rolón, G. Chavira, J. Orozco, J.P. Soto, Towards a framework for evaluating usability of business process models with BPMN in health sector, Procedia Manufacturing 3 (2015), 5603–5610.
[11] Medical University of Vienna, Hepatitis in der Schwangerschaft – Leitlinie, http://www.meduniwien.ac.at/frauenheilkunde/SOP/GH/Hepatitis_in_der_Schwangerschaft_3.pdf, last access: 1.1.2018.
[12] Activiti BPMN 2.0 Platform, http://activiti.org/, last access: 1.1.2018.
[13] T. Rademakers, Activiti in Action – Executable Business Processes in BPMN 2.0, Manning Publications, Shelter Island, 2012.
[14] G. Hripcsak, Writing Arden Syntax Medical Logic Modules, Comput Biol Med 24 (1994), 331–363.
[15] K.-P. Adlassnig, A. Rappelsberger, Medical knowledge packages and their integration into health-care information systems and the World Wide Web, Stud Health Technol Inform 136 (2008), 121–126.
[16] M. Samwald, K. Fehre, J. de Bruin, K.-P. Adlassnig, The Arden Syntax standard for clinical decision support: Experiences and directions, J Biomed Inform 45 (2012), 711–718.

[17] R.A. Jenders, P. Haug, K. Fehre, K.-P. Adlassnig, HL7 Arden Syntax Implementation Guide, Release 1, October 2014. Health Level Seven International, http://www.hl7.org/implement/standards/product_brief.cfm?product_id=371, last access: 1.1.2018.
[18] ARDENSUITE, http://www.medexter.com/products/ardensuite, last access: 1.1.2018.
[19] R.T. Fielding, Architectural Styles and the Design of Network-Based Software Architectures, Dissertation, University of California, Irvine, 2000.
[20] Schuh, J.S. de Bruin, W. Seeling, Clinical decision support systems at the Vienna General Hospital using Arden Syntax: Design, implementation, and integration, Artif Intell Med (2018), doi: 10.1016/j.artmed.2015.11.002.
[21] J.S. de Bruin, C. Schuh, W. Seeling, E. Luger, M. Gall, E. Hütterer, G. Kornek, B. Ludvik, F. Hoppichler, K. Schindler, Assessing the feasibility of a mobile health-supported clinical decision support system for nutritional triage in oncology outpatients using Arden Syntax, Artif Intell Med (2018), doi: 10.1016/j.artmed.2015.10.001.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-25

Going Mobile: An Empirical Model for Explaining Successful Information Logistics in Ward Rounds

Moritz ESDAR[a,1], Jan-David LIEBE[a], Birgit BABITSCH[b], Ursula HÜBNER[a]
[a]*Health Informatics Research Group, Osnabrück University AS, Germany*
[b]*Human Sciences, New Public Health, University Osnabrück, Germany*

Abstract. Background: Medical ward rounds are critical focal points of inpatient care that call for uniquely flexible solutions to provide clinical information at the bedside. While this fact is undoubted, adoption rates of mobile IT solutions remain rather low. Objectives: Our goal was to investigate if and how mobile IT solutions influence successful information provision at the bedside, i.e. clinical information logistics, as well as to shed light at socio-organizational factors that facilitate adoption rates from a user-centered perspective. Methods: Survey data were collected from 373 medical and nursing directors of German, Austrian and Swiss hospitals and analyzed using variance-based Structural Equation Modelling (SEM). Results: The adoption of mobile IT solutions explains large portions of clinical information logistics and is in itself associated with an organizational culture of innovation and end user participation. Conclusion: Results should encourage decision makers to understand mobility as a core constituent of information logistics and thus to promote close end-user participation as well as to work towards building a culture of innovation.

Keywords. mHealth, Organizational Culture, Diffusion of Innovation, Patients' Rooms, Clinical Rounds, Electronic Health Records

1. Introduction

Ward rounds are uniquely information intense workflows as they are arguably the most important focal points for medical decision making in secondary care [1,2]. This poses unique challenges to hospital IT-systems in terms of flexible and effective information provision in close proximity to the point of care. It also holds promises with respect to efficiency gains [3], enhanced access to information [4] and increasing quality of care [5]. Meeting key requirements of successful clinical information logistics [6] in ward rounds in terms of providing the right information for the right person, at the right time and in the right quality, requires innovative and mobile IT solutions to be applied. Although there has been rising adoption and increased attention in research towards mobile solutions following the society-wide uptake of tablet and smartphone usage, adoption rates in hospital settings remain surprisingly low [7]. This poses questions about their efficiency and usefulness in clinical environments. Benefits might be considerably hampered by negative externalities such as distractions from physician-patient interactions [8,9], usability issues [4] as well as security concerns that might outweigh potential benefits.

[1] Corresponding author, Moritz Esdar, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: m.esdar@hs-osnabrueck.de.

In accordance with modest adoption rates, workflow support through IT in ward rounds is often perceived significantly poorer compared to other clinical core processes such as clinical admission and pre- and post-surgery workflows or discharge [10]. Thus, we were interested in researching current adoption rates of mobile IT solutions in ward rounds across hospitals in the DACH-region (Germany-Austria-Switzerland) and in investigating the association between adoption rates and the perceived quality of workflow support from a user perspective. Successful clinical information logistics was thereby regarded as the manifestation of effective workflow support [11].

We furthermore aimed at gaining insights on how hospitals can bring innovative health IT such as mobile IT solutions into practice, i.e. what socio-organizational factors distinguish adopters from non-adopters. In times of increasingly fast innovation cycles in health IT, hospitals need to be able to flexibly adopt innovations within clinical workflows in order to be able to deliver high quality care and stay competitive [12]. Building on previous works we specifically considered the degree of user participation in the different stages of IT projects (in strategy development, implementation, evaluation etc.) and an innovation-friendly culture created by the top management to be possible antecedents of higher adoption rates of mobile IT solutions as both are regarded as key components of an organization's innovation capabilities [13-15]. The main goal therefore was to elaborate and test a generalizable framework from a user perspective that maps out preconditions and consequences of mobility and successful information logistics in ward rounds on a large scale.

2. Methods

2.1. Material

Data used in the analyses were obtained in the context of the international initiative "IT Report Healthcare" that aimed amongst others to assess the perceived IT-usage, IT-workflow-support, IT-quality and engagement of clinical staff in IT-projects. Most items were developed on the basis of existing surveys and scales [10,13,16] whereas scales for measuring information logistics in ward rounds as well as indicators of participation were newly developed and pretested in two iterations by a total of 14 experts (including health IT scientists, statisticians, management researchers, executive health professionals and one psychologist). The questionnaire targeted clinical users in German, Austrian and Swiss hospitals. In order to yield an overview of the entire hospital, the medical and nursing directors, as representatives of their hospital, were asked to provide answers that represent the prevailing view of the front-line clinicians. Data were collected via an online questionnaire between June and September 2017. We received fully completed responses from 373 out of a total of 2421 hospitals contacted (response rate: 15.4%). 85.2% of responses came from Germany, 7.0% from Austria and 7.8% from Switzerland.

2.2. Data Analysis

Since we aimed at statistically explaining both, antecedents and consequences of mobile IT solutions in order to work towards a comprehensive and more generalizable model, we applied structural equation modelling (SEM). SEM is the most common family of statistical methods available for testing complex cause-effect relationships, especially when latent variables (i.e. variables, that are not directly observed) are involved [17].

The model was specified to explain the variance in the endogenous variable "mobility" by the degrees of "innovation culture" and "participation" as well as to explain the variance in "clinical information logistics" by all other variables through direct and indirect effects, mediated by the degree of mobility (Fig. 1). All four variables were specified as latent variables, reflected by their respective indicators from the survey (Tab. 1). Thus, four reflective measurement models were combined in one structural model.

SEM offers a variety of possible techniques and algorithms with covariance-based SEM (CB-SEM) techniques being the most common [18]. CB-SEM has a primarily confirmatory character in that it focusses on model-fit and is therefore mostly used to test and confirm existing models within theory development. In contrast to CB-SEM, variance-based SEM, known as partial least squares (PLS-SEM) has a stronger focus on predicting target constructs and on the identification of key drivers with a more exploratory character [19]. However, in recent years a modified algorithm based on PLS has been developed that produces results largely consistent with CB-SEM, called consistent partial least squares (cPLS-SEM) which is increasingly gaining popularity [20]. We decided to apply cPLS-SEM to our data since it is also known to be advantageous in that it does not require multivariate normality and tolerates ordinal scaled data [21].

In order to assess the reflective models, the latent constructs "innovation culture", "participation", "mobility" and "clinical information logistics" were tested for reliability as well as convergent and discriminant validity using Cronbach's α, congeneric reliability (CR), the average variance extracted (AVE) and heterotrait-monotrait ratios of correlations (HTMT) [21]. Relations within the structural model were investigated based on the direct, total and indirect effects with p-values and confidence intervals obtained from 1,000 bootstrap replications. All analyses were performed using *SmartPLS (v. 3.2.7)*.

3. Results

The responses from our participants show that the adoption rates of mobile IT solutions in hospital still seem to have room for improvement (Tab. 1) – especially in Germany. On average, respondents indicated about 50% of the clinical units to have Wi-Fi available (GER: $\bar{x} = 45.7\% \pm 41.7\%$; AUT: $\bar{x} = 66.2\% \pm 40.1\%$; CH: $\bar{x} = 80.0\% \pm 34.2\%$) and that 45% of hospital wards have mobile access to patient data (GER: $\bar{x} = 41.7\% \pm 42.3\%$; AUT: $\bar{x} = 64.0\% \pm 41.2\%$; CH: $\bar{x} = 77.6\% \pm 37.7\%$) - although on average only 3.54 (out of 10) patient data types[1] (GER: $\bar{x} = 3.3 \pm 4.0$; AUT: $\bar{x} = 4.4 \pm 3.9$; CH: $\bar{x} = 5.5 \pm 4.4$) were reported to be accessible on mobile devices. About 20% of the surveyed hospitals reported to have at least eight data types available and therefore seemed to have implemented more comprehensive solutions.

[1] Patient identity data, case data (diagnosis and therapy codes), orders, results (text), results (images), results (electrophysiology), kardex with medication and vital signs, warnings, check lists, others

Table 1. Indicators & descriptive results (n = 373). [a] Likert-Scale (1 = "strongly disagree, 2 = "disagree", 3 = "neutral", 4 = "agree", 5 = "strongly agree"). [b] 1 = "no participation at all" … 10 = intensive participation". [c] composite scores, ranging from 0-10.

Code	Question	Sub-question	$\bar{x}$	SD
IC1	Please indicate your (dis-) agreement with the following statements.[a]	"Our executive board actively promotes the initiation of innovative IT projects."	3.54	1.13
IC2		"Our executive board regularly perceives IT as a mere expense factor."	2.60	1.12
IC3		"Our hospital shows great agility and flexibility when it comes to implementing new IT solutions."	3.08	1.17
IC4		"Our hospital has a well-defined future vision that is shared by the IT department."	3.33	1.19
IC5		"We usually take IT into account when working on new medical or nursing related projects."	3.41	1.06
IC6		"There is a culture of tolerance in our hospital when dealing with mistakes and failing projects."	3.48	0.93
IC7		"Our IT is capable to react quickly in face of changing requirements."	2.77	1.08
IC8		"We openly communicate and discuss new IT projects in our hospital among all involved staff."	3.22	1.08
IC9		"Our executive board explicitly demands ideas and suggestions on how to innovate our IT."	3.04	1.09
Par1	Please indicate the degree of participation of clinical staff (e.g. physicians or nurses) in issues concerning your hospital information system. Degree of participation…[b]	…in strategy development?	5.39	2.53
Par2		…in identifying and defining clinical requirements for IT applications?	5.60	2.48
Par3		…in evaluating and selecting IT applications?	5.05	2.60
Par4		…in the implementation process?	5.88	2.64
Par5		…in developing and conducting teaching and training during the implementation phase?	6.04	2.73
Par6		…in teaching and training of new staff after implementation?	6.01	2.78
Par7		…in evaluating and modifying IT applications?	5.20	2.77
Mob1	What patient data are accessible on mobile devices during ward rounds?[c]		3.54	4.12
Mob2	Which devices (hardware) are available for documenting and processing patient data?[c]		2.75	2.67
Mob3	In how many clinical units (in percent) is Wi-Fi available (Wi-Fi coverage)?		49.83	4.22
Mob4	How many wards (in percent) have mobile electronic access to their patient data?		44.80	4.30
CIL1	Please evaluate the quality of electronic information provision during ward rounds. The required clinical information is…[a]	…available at the right place.	3.31	1.51
CIL2		…available for the right person.	3.71	1.38
CIL3		…correct and complete.	3.51	1.31
CIL4		…legible and clear.	3.87	1.33
CIL5		…up-to-date.	3.86	1.28
CIL6		…being made available in a timely manner.	3.68	1.31
CIL7		…provided in a user-friendly manner.	3.17	1.38
CIL8		It takes a reasonable time to compile the relevant information.	3.22	1.35

Construct reliability and validity of the four specified latent variables proved to be overall satisfactory with reliability measures (Cronbach's α & congeneric reliability) well above

the common thresholds and with more than 50% average variance extracted from the reflective indicators (Tab. 2) [22]. Also, heterotrait-monotrait ratios of correlations between all latent variables fell well below 0.85 and therefore indicated sufficient discriminant validity between the latent variables [17]. Relatively high factor loadings in all measurement models underpinned their validity (Fig. 1). Whereas values for "participation", "mobility" and " clinical information logistics" were very good, the construct validity and reliability of "innovation culture" showed to be lower, yet measures lay above common thresholds of acceptance [21]. The specified reflective constructs can thus be confirmed in our data.

Table 2. Construct reliability and validity (n = 373)

Latent Variable	Cronbach's α	Congeneric Reliability	Average Variance Extracted
Innovation culture	.76	.82	.51
Participation	.94	.94	.70
Mobility	.93	.93	.76
Clinical information logistics	.96	.96	.75

Figure 1 displays significant direct effects between all domains except for the prediction of "clinical information logistics" by "participation", which only showed an indirect significant effect through the degree of "mobility" (Tab. 3). The strongest associations appeared between "mobility" and "clinical information logistics" with a ß-coefficient of .45 and between "innovation culture" and "clinical information logistics" (ß = .40) whereas the latter effect was mostly mediated by "mobility".

Table 3. Total and indirect effects with bias corrected 95% confidence intervals (CI) from bootstrapping. Legend: IC – innovation culture, MOB – mobility, PART – participation, CIL – clinical information logistics

			95% CI		
	Path	**Coefficient**	**Lower**	**Upper**	**p-value**
Total Effects	IC → MOB	.32	.15	.52	< .001
	PART → MOB	.24	.05	.41	< .01
	MOB → CIL	.45	.36	.53	< .001
	IC → CIL	.40	.21	.54	< .001
	PART → CIL	.23	.07	.38	< .01
Indirect Effects	PART → MOB → CIL	.11	.02	.19	< .01
	CIC → MOB → CIL	.15	.07	.25	< .001

Almost one third of the variance in "mobility" can be explained solely by "innovation culture" and the degree of "participation". About half of the variation in "clinical information logistics" could be explained by all other latent variables, mostly by the degree of "mobility" (Fig. 1). The standardized root mean residual (SRMR) value of 0.05 indicated furthermore a satisfactory model fit to the data [23].

Despite considerable differences in the adoption rates of mobile IT solutions across countries, the model parameter did not change when excluding the Austrian and Swiss hospitals from the sample. Calculating the model with data only from Austria and Switzerland without Germany was not warranted due to insufficient sample sizes (n(AUT) = 26; n(CH) = 29) [24].

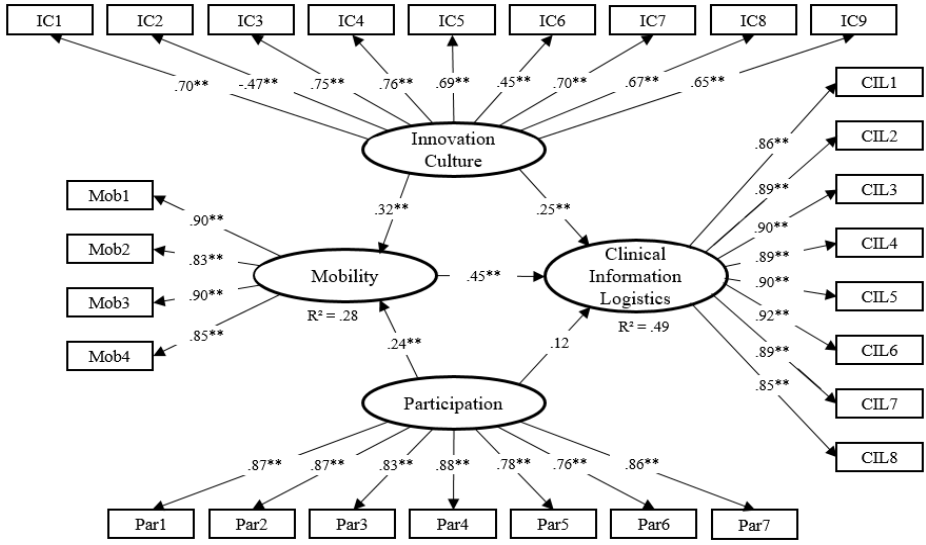


Figure 1. Structural model (n = 373, standardized root mean residual (SRMR) = .05; **p<.01)

4. Discussion

We found adoption rates of mobile IT solutions in ward rounds to still be on a rather low level across hospitals in the DACH-region. While on average, about 45% of hospital wards potentially seem to have mobile access to patient data, not all patient data are available electronically, suggesting that many hospitals still use a combined approach of paper based and electronic solutions. The widespread availability of mobile devices also seems to be calling for improvements. Adoption rates of mobile IT solutions appear to be higher in Austrian hospitals and especially in Switzerland where respondents reported that on average 78% of the hospital wards have mobile access to patient data. Those differences go hand in hand with results from international eHealth-benchmarks that report comparable differences between the three countries [25-27].

Against the background of these deficits, the question arises as to how improvements can be made possible. The results of this structural model provide a contribution to better understand what mobility actually is and how it can be facilitated – from a users' perspective. First, it clearly demonstrates that the implementation of mobile IT solutions in hospitals is not an end in itself but determines successful information provision, i.e. "clinical information logistics", which is one of the grand goals of health IT in itself. "Clinical information logistics", the more abstract umbrella term, can be tied to the presence of mobile IT solutions (ß = .45), something that every user can experience. Mobility thus is to a significant part the tangible manifestation of "clinical information logistics" in ward rounds. Second, "mobility" in combination with "innovation culture" can explain 49% of the variability in "clinical information logistics", which is quite a considerable proportion given that only the users' perspective was taken into account. This finding demonstrates a powerful mechanism, i.e. innovation culture pervading the organization, to explain "clinical information logistics". Third, this powerful force is an enabler of "mobility" itself. Together with "participation", it seems to drive "mobility". Both concepts are at work when mobility has to be achieved in that innovation culture

initiates and shapes the necessary steps by the organization's spirit in a top-down fashion and is then ideally accompanied by comprehensive end-user involvement from the bottom up. Mobile ward round scenarios closely touch on the clinicians' daily work practice. It therefore seems plausible that close participation of end users across different stages of implementation is positively associated with higher degrees of mobility. The positive influence of both factors corresponds to similar findings that emphasize the importance of organizational culture and user involvement for successful implementation of health information technologies as such [14,28]. This study specifies these general findings focusing on the crucial areas of mobility and clinical information logistics and thus eventually on information continuity. Creating "mobility" itself requires "participation" and an organizational "innovation culture".

According to the measures of reliability and factor loadings we can overall confirm that the hypothesized constructs (i.e. latent variables) seem to be well reflected by their respective indicators. This, together with a satisfactory model fit and acceptable discriminant validity measures, suggests that there are no crucial misspecifications in our model. Although "participation" was operationalized differently in preceding works [13], this factor again demonstrates its internal consistency and substantial relevance. In contrast, reliability measures of "innovation culture" might be improved by removing some variables.

Given that this study represents the users' perspective some portions of the construct "mobility" remain unexplained with the chosen model specifications which opens the door to other possible factors to consider. They include variables of the organization and its approach to deploy and maintain IT technology such as professionalism of management structures or processes, financial power, IT-service quality, legal regulations and health IT vendors [14]. Demographic covariates (i.e. hospital size, ownership and teaching status) are –amongst others– known to be influencing health IT adoption rates [12,29], but lay beyond the scope of this model. Those factors could be accounted for in future model extensions that look beyond organizational "innovation capabilities". The main limitation stems from a modest response rate that might have caused a self-selection bias. However, the relatively large sample size mitigates that effect.

While adoption rates of mobile IT solutions in German, Austrian and Swiss hospitals are yet to be improved, this study is one of the first to empirically demonstrate the clear connection between mobility and clinical information logistics and provides evidence that an organizations' innovation culture and the degree of participation seem to drive mobile health IT innovations. The results should encourage decision makers including chief information officers to promote close participation of end users in all phases of project management as well as to work towards building a culture of innovation.

Acknowledgement

This research was funded by the Federal State of Lower Saxony Germany (VWZN3062).

References

[1] P.H. Pucher, R. Aggarwal, A. Darzi, Surgical ward round quality and impact on variable patient outcomes, Ann. Surg. 259 (2014), 222–6.

[2] C. Nikendei, B. Kraus, M. Schrauth, S. Briem, J. Jünger, Ward rounds: how prepared are future doctors? Med. Teach. 30 (2008), 88–91.
[3] R. Fleischmann, J. Duhm, H. Hupperts, S.A. Brandt, Tablet computers with mobile electronic medical records enhance clinical routine and promote bedside time: a controlled prospective crossover study, J. Neurol. 262 (2015); 532–40.
[4] E.C. Lehnbom, K. Adams, R.O. Day, J.I. Westbrook, M.T. Baysari. iPad use during ward rounds: an observational study, Stud. Health. Technol. Inform., 204 (2014), 67–73.
[5] M. Prgomet, A. Georgiou, J.I. Westbrook. The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review, JAMIA, 16 (2009), 792–801.
[6] M. Esdar, U. Hübner, J.D. Liebe, J. Hüsers, J. Thye, Understanding latent structures of clinical information logistics: A bottom-up approach for model building and validating the workflow composite score, Int. J. Med. Inform. 97 (2017), 210–220.
[7] J. Sclafani, T.F. Tirrell, O.I. Franko, Mobile Tablet Use among Academic Physicians and Trainees. J. Med. Syst. 37 (2013), 9903.
[8] M.T. Baysari, K. Adams, E.C. Lehnbom, J.I. Westbrook, R.O. Day, iPad use at the bedside: a tool for engaging patients in care processes during ward rounds? Intern. Med. 44 (2014), 986–90.
[9] O.A. Alsos, A. Das, D. Svanæs, Mobile health IT: the effect of user interface and form factor on doctor-patient communication. Int. J. Med. Inform. 81 (2012), 12–28.
[10] J.D. Liebe, U. Hübner, M.C. Straede, J. Thye, Developing a Workflow Composite Score to Measure Clinical Information Logistics. A Top-down Approach. Methods Inf. Med. 54 (2015), 424–33.
[11] A. Winter, R. Haux , E. Ammenwerth, B. Brigl, N. Hellrung, F. Jahn F, Health Information Systems, Springer, London, 2011.
[12] D.E. Leidner, D. Preston, D. Chen, An examination of the antecedents and consequences of organizational IT innovation in hospitals. J. Strateg. Inf. Syst. 19 (2010), 154–70.
[13] M. Esdar, J.D. Liebe, J.P. Weiß, U. Hübner. Exploring Innovation Capabilities of Hospital CIOs: An Empirical Assessment. Stud. Health. Technol. Inform. 235 (2017), 383–7.
[14] K. Cresswell, A. Sheikh, Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review, Int. J. Med. Inform. 82 (2013), e73-86.
[15] R. Ford, Complex leadership competency in health care: towards framing a theory of practice, Health Serv. Manag. Res. 22 (2009), 101–14.
[16] A.K. Jha, C.M. DesRoches, E.G. Campbell, K. Donelan, S.R. Rao, T.G. Ferris et al. Use of Electronic Health Records in U.S. Hospitals, N Engl J Med 360 (2009), 1628–38.
[17] T.D. Little, R.B. Kline, editors, Principles and practice of structural equation modeling. The Guilford Press, New York, 2016.
[18] J.F. Hair, M. Gabriel, V. Patel, AMOS Covariance-Based Structural Equation Modeling (CB-SEM): Guidelines on Its Application as a Marketing Research Tool, Brazilian Journal of Marketing 13 (2014).
[19] J.F. Hair, C.M. Ringle, M. Sarstedt, PLS-SEM: Indeed a Silver Bullet, JMTP 19 (2011), 139–152.
[20] TK Dijkstra, J. Henseler, Consistent partial least squares path modeling, MIS Quarterly 39 (2015), 297.
[21] J.F. Hair, G.T. Hult, C.M. Ringle, M. Sarstedt, A primer on partial least squares structural equation modeling (PLS-SEM). Sage Publ., Los Angeles 2014.
[22] C. Fornell, D.F. Larcker, Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. J. Mark. Res. 18 (1981), 39.
[23] L. Hu, P.M. Bentler, Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification, Psychol. Methods 3 (1998), 424–53.
[24] E.J. Wolf, K.M. Harrington, S.L. Clark, M.W. Miller, Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety, Educ. Psychol. Meas 76 (2013), 913–34.
[25] P. Arak, A. Wójcik, Transforming eHealth into a political and economic advantage, Polityka Insight, Warszawa, 2017.
[26] J. Hüsers J, U. Hübner, M. Esdar, E. Ammenwerth, W.O. Hackl, L. Naumann, J.D. Liebe, Innovative Power of Health Care Organisations Affects IT Adoption: A bi-National Health IT Benchmark Comparing Austria and Germany. J. Med. Syst. 41 (2017), 33.
[27] R. Sabes-Figuera, European Hospital Survey: Benchmarking deployment of e-Health services (2012–2013) – Composite Indicators on eHealth Deployment and on Availability and Use of eHealth functionalities, JRC Working Papers.
[28] A. M. Høstgaard, P. Bertelsen, C. Nøhr. Methods to identify, study and understand End-user participation in HIT development, BMC Med. Inform. Decis. Mak. 11 (2011), 57.
[29] C.M. DesRoches, D. Charles, M.F. Furukawa, M.S. Joshi, P. Kralovec, F. Mostashari et al., Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. Health Aff (Millwood) 32 (2013), 1478–85.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-33

Use of Mobile Apps Among Medical and Nursing Students in Iran

Abbas SHEIKHTAHERI[a,b], Farzaneh KERMANI[a,1]
[a] *Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran*
[b] *Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran*

Abstract. Mobile technologies have a positive impact on patient care and cause to improved decision making, reduced medical errors and improved communication in care team. The purpose of this study was to investigate the use of mobile technologies by medical and nursing students and their tendency in future. This study was conducted among 372 medical and nursing students of Tehran University of Medical Science. Respectively, 60.8% and 62.4% of medical and nursing students use smartphone. The most commonly used apps among medical students were medical dictionary, drug apps, medical calculators and anatomical atlases and among nursing students were medical dictionary, anatomical atlases and nursing care guides. Also, the use of decision support systems, remote monitoring, patient imagery and remote diagnosis, patient records documentation, diagnostic guidelines and laboratory tests will be increased in the future.

Keywords. Mobile technology, mobile education, medical education, nursing education

1. Introduction

The use of Smartphone, personal digital assistant (PDA) and other mobile and handheld devices with immediate access to health information has had a positive impact on patient care. In addition, these technologies led to improved decision-making process and reduced medical errors, improved communication between the treatment team and enhanced telemedicine capabilities. Currently, doctors and nurses are able to store textbooks and video tutorials, and use sources such as medical calculators [1], diagnostic guidelines, management and drug references apps [2] on their Smartphone. Additionally, library and educational apps [3], simulated health care environment [4], laboratory test and drug interaction guides, nursing care and clinical examination guidelines [5] are also available features for mobile devices. The use of pharmaceutical and therapeutic information on Smartphone is effective in providing patient care reports [6], better communication between providers [7] and increasing information exchange in health care settings [8]. A study in the USA showed that doctors believed that the use of PDAs in the field of pharmaceutical and therapeutic information was effective in providing patient care reports in 91% of cases [6].

In addition to clinical usage, the educational use of this technology is highly prevalent amongst students. A study in the USA showed that more than 70% of medical

[1] Corresponding author, Farzaneh Kermani, PhD candidate in Medical Informatics, Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran; Email: Kermani-f@tak.iums.ac.ir.

students use these devices mostly for electronic reference books, medical databases, medical calendars, and patients tracking [9]. The study of medical students and junior doctors in United Kingdom also showed that most of these people have between 1 to 5 medical apps on their phone [1]. Another study showed that 98% of nursing students use drug guidelines and 83% of them use medical dictionary on these devices [2]. Students can use apps for predicting, checking for drug interactions and counseling on differential diagnosis [9] that lead to improved learning, reduced medical errors, increased accuracy of students at the clinic, and saving training time [10].

Although there are many studies in other countries, there is not much study in Iran about the use of this technology among medical and nursing students. Given the fact that these two groups form the next generation of physicians and nurses, one can expect the use of mobile technology to increase among new-generation students. However, there is not much information about the use of this type of technology and its type of use among this group of students in Iran. Therefore, the purpose of this study was to investigate the use of mobile technologies by medical and nursing students and their tendency in using these technologies in future.

2. Method

This research was conducted in teaching hospitals of Tehran University of Medical Sciences in 2016. The research population included medical and nursing students working in these hospitals. The sample included 400 individuals who were invited to participate using simple random sampling and Morgan sampling table (200 individuals from every group). A questionnaire was used to collect data. The questionnaire included questions about demographic data, the current use and app of mobile technology, and favorite apps for future use. Validity of the questionnaire was evaluated using the comments of three faculty members of Health Information Technology and six medical and nursing students (three individuals from each group) to present their views on the transparency and necessity of the questions. The revised questionnaire was re-submitted to the same faculty members, and changes were made based on their comments.

In order to control the reliability, the questionnaire was first provided to 30 students (15 individuals from each group) and the amount of Kuder Richardson's alpha was calculated. The test-retest method was also used. The questionnaire was provided to 10 medical and nursing students in two phases in a 10-day interval and the correlation of the responses was calculated. No significant difference was observed in two phases. After collecting the questionnaires, the data analysis was performed using SPSS 16 and descriptive methods (frequency and percentage). Chi-square, Fisher test and t-test were used for comparing medical and nursing students.

3. Results

A total of 372 students (194 medical students, 178 nursing students) participated in this study. Female subjects accounted for the majority of the participants (51.9%) and most of students (83.3%) were under 30 years of age. The average hospital work experience of medical and nursing students was 42.5 ± 4.2 and 30.5 ± 2.9 months, respectively. The most used technologies among medical and nursing students were Smartphone (60.8% vs. 62.4%, respectively). Medical and nursing students had 41.4 ± 21.02 and 29.6 ± 21.59

months' experience in using mobile technology ($p<0.001$). Medical and nursing students had had used these technologies in clinical and educational tasks for 28.9 ± 17.2 and 18.7 ± 15.4 months, respectively ($p<0.001$) and only 6.5% of medical students and 18.9% of nursing students stated that they rarely use or do not use the mobile technology in clinical and educational tasks.

Most students used this technology more than once or twice a day. The smartphone was the most favorite device for future use among medical and nursing students with 62.4% and 59.7%, compared to other technologies like tablets. Also, 88.2% and 91.8% of medical and nursing students, respectively, were willing to use the mobile technology in the future.

The current use of most apps was less than 40% in both groups (Tables 1 and 2). There was no significant difference between the two groups in terms of using most of these apps. The medical students used apps such as medical dictionary, drug apps, medical calculators, medical video tutorials, decision support and access to databases more than nurses ($P<0.05$). Nurses used nursing care guides more than medical students ($P<0.05$).

Considering the willingness of students to use these apps in the future, both groups showed willingness rate of less than 40%. In addition, there was no significant difference between the two groups in terms of using most of these apps in the future, but nursing students would be more in need of medical dictionary, nursing care guides, anatomical atlases, educational management apps than medical students (Tables 1 and 2, sorted alphabetically).

According to Tables 1 and 2, there will be an increase in demand for most of these apps in the future, with the highest increase for apps used in decision support (31.7% for medical students and 18.2% for nursing students), monitoring and remote care (26.8% for medical students and 20% for nursing students), patient imagery and remote diagnosis (23.7% in medical students and 12.9% in nursing students), patient record

Table1. Comparison of usage rates of apps (current and future) among medical students (N=186)

	Current use n (%)	**Tendency for future use** n(%)	**Change rate** %
Anatomical atlases	81(43.5)	56(30.1)	-13.4↓
Calendar (visit schedule)	19(10.2)	41(22)	+11.8↑
Clinical guidelines	39(21)	55(29.6)	+8.6↑
Decision support systems	4(2.2)	63(33.9)	+31.7↑
Diagnostic guide & laboratory tests	29(15.6)	59(31.7)	+16.1↑
Drug apps (doses & drug interactions)	112(60.2)	72(38.7)	-21.5↓
Educational management apps	14(7.5)	20(10.8)	+3.3↑
Electronic reference books	60(32.3)	48(25.8)	-6.5↓
Identification Program (student card, library)	6(3.2)	10(5.4)	+2.2↑
Library apps (access to digital library, book depository)	10(5.4)	27(14.5)	+9.1↑
Medical calculators	85(45.7)	45(42.2)	-3.5↓
Medical database (Medline)	61(32.8)	58(31.2)	-1.6↓
Medical dictionary[1]	132(71)	78(41.9)	-29.1↓
Medical encyclopedia[2]	43(23.1)	56(30.1)	+7↑
Medical news track	49(26.3)	49(26.3)	0
Medical video tutorials	50(26.9)	51(27.4)	+0.5↑
Nursing care guidelines	14(7.5)	23(12.4)	+4.9↑
patient imagery and remote diagnosis	11(5.9)	55(29.6)	+23.7↑
Patient records documentation	9(4.8)	39(21)	+16.2↑
Physical examinations apps	18(9.7)	47(25.3)	+15.6↑
Remote monitoring & patient care	20(10.8)	70(37.6)	+26.8↑

[1] Medical dictionary: a lexicon for words used in medicine

[2] Medical encyclopedia: a comprehensive written compendium that holds information about diseases, medical conditions, tests, symptoms, injuries, and surgeries.

Table2. Comparison of usage rates of apps (current and future) among nursing students (N=170)

	Current use n(%)	**Tendency for future use** n(%)	**Change rate** %
Anatomical atlases	61(35.9)	75(44.1)	+8.2↑
Calendar (visit schedule)	17(10)	37(21.8)	+11.8↑
Clinical guidelines	31(18.2)	57(33.5)	+15.3↑
Decision support systems	12(7.1)	43(25.3)	+18.2↑
Diagnostic guide & laboratory tests	37(21.8)	65(38.2)	+16.4↑
Drug apps (doses & drug interactions)	82(48.2)	57(33.5)	-14.7↓
Educational management apps	19(11.2)	40(23.5)	+12.3↑
Electronic reference books	41(24.1)	44(25.9)	+1.8↑
Identification Program (student card, library)	12(7.1)	28(16.5)	+9.4↑
Library apps (access to digital library, book depository)	14(8.2)	36(21.2)	+13↑
Medical calculators	22(12.9)	35(20.6)	+7.7↑
Medical database (Medline)	20(11.8)	48(28.2)	+16.4↑
Medical dictionary	89(52.4)	95(55.9)	+3.5↑
Medical encyclopedia	37(21.8)	58(34.1)	+12.6↑
Medical news track	35(20.6)	58(34.1)	+13.5↑
Medical video tutorials	30(17.6)	51(30)	+12.4↑
Nursing care guidelines	44(25.9)	82(48.2)	+22.3↑
patient imagery and remote diagnosis	11(9.5)	38(22.4)	+12.9↑
Patient records documentation	13(7.6)	41(24.1)	+16.5↑
Physical examinations apps	22(12.9)	46(27.1)	+14.2↑
Remote monitoring & patient care	15(8.8)	49(28.8)	+20↑

documentation and access to the records (16.2% among medical students and 16.5% among nursing students), diagnostic guidelines and laboratory tests (16.1% for medical students and 16.4% for nursing students). In comparison, the rates of students' use of some apps, such as medical dictionary (29.1% among medical students), medical calculators (3.5% among medical students) and anatomical atlas (13.4% among medical students) will be decreased. Figure 1 shows the current and future use of apps among all students.

4. Discussion

Newzoo's 2017 Global Mobile Market Report shows the top 50 countries in terms of smartphone users in 2017. According to this report, Iran is among the top 50 countries by 30 million smartphone users. China, India and United States respectively have the highest rates (717.31, 300.12, 226.29 million smartphone users). [11] Smartphones are also the most common technology used by students. According to a review study, about 60% -70% of students and physicians used PDA in educational and clinical tasks [9], which is consistent with the results of this study. A study in the USA showed that 96% of nursing students used mobile technology in clinical tasks [2]. In the UK, 78.3% and 39.9% of medical students used smartphones for educational and clinical purposes, respectively [1]. According to these studies, the use of mobile technology, especially the smartphone, is increasing among medical and nursing students, which is consistent with the results of the present study. Most medical and nursing students used this technology several times a day for clinical and educational use, indicating that the daily use of mobile technology is relatively high. In the United Kingdom, 47% of medical students used their device at home or at point of care at least once a week [12]. In Sweden, this rate was 1 to 5 times a day for nursing students and up to 10 times a day [13]. These studies show a high rate of daily use of mobile technologies among students, which is roughly consistent with our results.

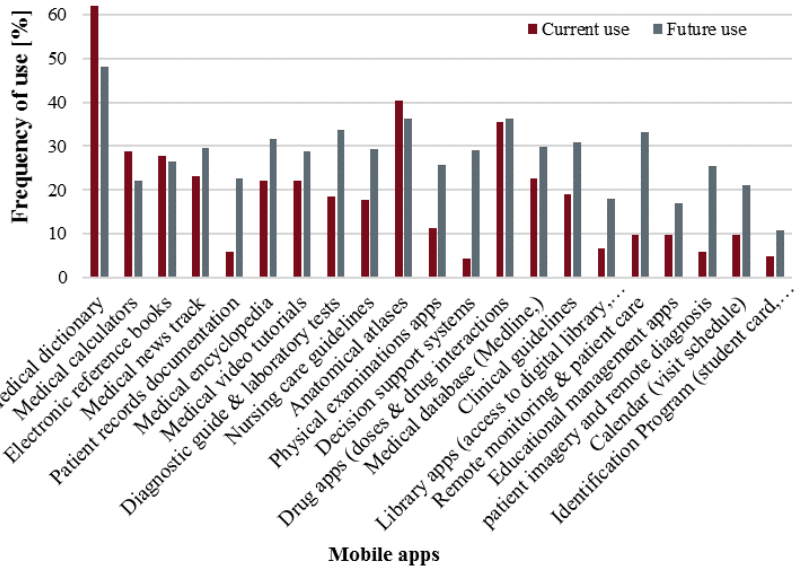


Figure 1. Comparison of usage rates of apps (current and future) among medical and nursing students

The results showed that the most commonly apps used by medical students included medical dictionary, drug apps, medical calculators and anatomy atlases. In this regard, according to a review study, the most commonly apps used by students included drug information apps, medication references, medical calculators and information management [9]. A study in the UK has reported that the major use of apps among medical students is related to drug handbook and medical books [12]. These results are roughly similar to our results. The apps widely used by nursing students included medical dictionary, drug apps, anatomy atlas and nursing care guides. A study in the USA showed that the most desired apps by nursing students include clinical guidelines and dictionary [2]. A review study also showed that most nursing students used these apps to receive reference information [14], which is almost consistent with our results.

Future apps needed by medical students will include medical calculators, medical dictionary, drug apps and remote monitoring and patient care. Nursing students will also need medical dictionary, nursing care guides, anatomy atlases, diagnoses and laboratory test guidelines. In Germany, medical students need some apps to search for diagnostic, therapeutic and predictive information, access to patient records, interaction and social communication, organization of medical education, remote patient monitoring and online reporting [15]. As for nursing students, other apps like access to patient information, tests and reference values of tests, knowledge resources such as reference books, information about diseases, preparation for patient examination, medical calculators and the possibility of taking notes from journals were reported [16]. According to these results, current and future needs of Iranian students are similar to those of other developed countries. However, these requirements are also subject to the conditions of that country, the relevant educational programs and the special educational needs of the students of each country.

According to the current study, demand for most of these mobile apps will increase in the future. The highest increase is related to remote monitoring and patient care remote diagnosis, patient records documentation, diagnostic guidelines and laboratory tests.

There will a reduction in the use of some apps, such as the medical dictionary, medical calculators and anatomy atlas. In this regard, there is no report in other studies. In short, the use of mobile technology in both groups is beneficial for educational and clinical tasks, and students tend to use different types of apps on their mobile technologies, and there is increased tendency with regard to some apps. Therefore, university officials and app developers should consider developing new apps with multiple features to meet the growing needs. The results of this research can be used for developing appropriate mobile applications that are more interesting for students. In fact, app developers, especially those who are working in the education field may use these results to develop in-demand apps for future medical and nursing students. Some limitations should be considered. Although, the sample size is high, this study was conducted in one university. This university is the biggest medical university in the country with students from the whole country, however, we could not generalize the results to the whole country.

Acknowledgment

This study is a part of a research financially supported by Tehran University of Medical Sciences (Code 94-01-76-28289)

5. References

[1] K.F.B. Payne, H. Wharrad, K. Watts, Smartphone and medical related app use among medical students and junior doctors in the United Kingdom (UK): A regional survey, *BMC Med Inform Decis Mak* **12**(1) (2012), 1-11.

[2] L.E. George, L.J. Davidson, C.P. Serapiglia, S. Barla, A Thotakura, Technology in nursing education: A study of PDA use by students, *J Prof Nurs* **26**(6) (2010), 371-376.

[3] D.H. Shin, Y.J. Shin, H. Choo, K. Beom, Smartphones as smart pedagogical tools: Implications for smartphones as u-learning devices, *Comput Human Behav* **27**(6) (2011), 2207-2214.

[4] C.P. Hawkes, B.H Walsh, C.A. Ryan, E.M. Dempsey, Smartphone technology enhances newborn intubation knowledge and performance amongst paediatric trainees, *Resuscitation* **84**(2) (2013), 223-226.

[5] O.I. Franko. Smartphone apps for orthopaedic surgeons, *Clin Orthop Relat Res* **469**(7) (2011), 2042-2048.

[6] C.R. Dee, M. Teolis, A.D. Todd, Physicians' use of the personal digital assistant (PDA) in clinical decision making, *J Med Libr Assoc* **93**(4) (2005), 480-486.

[7] E. Ozdalga, A. Ozdalga, N. Ahuja, The smartphone in medicine: A review of current and potential use among physicians and students, *J Med Internet Res* **14**(5) (2012), 1-19.

[8] G.J. Putzer, Y. Park, The effects of innovation factors on smartphone adoption among nurses in community hospitals, *Perspect Health Inf Manag* **7** (2010), 1-18.

[9] A. Kho, L.E. Henderson, D.D. Dressler, S. Kripalani, Use of handheld computers in medical education, J *Gen Intern Medi* **21**(5) (2006), 531-537.

[10] A. Jotkowitz, J. Oh, C. Tu, D. Elkin, L.A. Pollack, H. Kerpen, The use of personal digital assistants among medical residents, *Med Teach* **28**(4) (2006), 382-384.

[11] Newzoo, Top 50 Countries by Smartphone Users and Penetration, available from: https://newzoo.com/insights/rankings/top-50-countries-by-smartphone-penetration-and-users/.

[12] B.S. Davies, J. Rafique, T.R. Vincent, J. Fairclough, M.H. Packer, R. Vincent, I. Haq, Mobile Medical Education (MoMEd)-how mobile information resources contribute to learning for undergraduate clinical students-a mixed methods study, *BMC Med Educ* **12**(1) *(*2012) 1-11.

[13] P.E. Johansson, G.I. Petersson, G.C. Nilsson, Nursing students' experience of using a personal digital assistant (PDA) in clinical practice—an intervention study, *Nurse Educ Today* **33**(10) (2013), 1246-1251.

[14] C. Strandell-Laine, M. Stolt, H. Leino-Kilpi, M. Saarikoski, Use of mobile devices in nursing student–nurse teacher cooperation during the clinical practicum: An integrative review, *Nurse Educ Today* **35**(3) (2015), 493-499.
[15] M. Sandholzer, T. Deutsch, T. Frese. A. Winter, Medical students' attitudes and wishes towards extending an educational general practice app to be suitable for practice: A cross-sectional survey from Leipzig, Germany, *Eur J Gen Pract* **22**(2) (2016), 141-146.
[16] M. Berglund, C. Nilsson, P. Révay, G. Petersson, G. Nilsson, Nurses' and nurse students' demands of functions and usability in a PDA, *Int J Med Inform* **76**(7) (2007), 530-537.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-40

Hospital CEOs Need Health IT Knowledge and Trust in CIOs: Insights from a Qualitative Study

Johannes THYE[a,1], Ursula HÜBNER[a], Jan-Patrick WEIß[a], Frank TEUTEBERG[b], Jens HÜSERS[a], Jan-David LIEBE[a] and Birgit BABITSCH[c]
[a]*Health Informatics Research Group, Hochschule Osnabrück, Germany*
[b]*Research Group Accounting & Information Systems, University Osnabrück, Germany*
[c]*Human Sciences, New Public Health, University Osnabrück, Germany*

Abstract. Background: IT is getting an increasing importance in hospitals. In this context, major IT decisions are often made by CEOs who are not necessarily IT experts. Objectives: Therefore, this study aimed at a) exploring different types of IT decision makers at CEO level, b) identifying hypotheses if trust exists between these different types of CEOs and their CIOs and c) building hypotheses on potential consequences regarding risk taking and innovation. Methods: To this end, 14 qualitative interviews with German hospital CEOs were conducted to explore the research questions. Results: The study revealed three major types: IT savvy CEOs, IT enthusiastic CEOs and IT indifferent CEOs. Depending on these types, their relationship with the CIO varied in terms of trust and common language. In case of IT indifferent CEOs, a potential vicious circle of lack of IT knowledge, missing trust, low willingness to take risks and low innovation power could be identified. Conclusion: In order to break of this circle, CEOs seem to need more IT knowledge and / or greater trust in their CIO.

Keywords. CEO, IT knowledge, IT decision making, CEO-CIO relationship

1. Introduction

IT in hospitals has an increasing impact on administrative and clinical processes [1,2]. Therefore, decision making related to hospital IT investments is a strategic task and a matter of the top management team, to increase hospital success. As has been shown, the following different styles of decision making processes can lead to final IT decisions: supported decisions, shared decisions and corporate decisions [3]. In the first case, the chief executive officer (CEO) makes the final decision after being supported by the chief information officer (CIO)[2]. In the second case, a team embracing the CEO, the CIO and clinicians try to build consensus and share the decision-making responsibilities. Finally, in the last case, the IT decision has been already made at corporate level [3]. Depending on the hospital CEO's IT knowledge and background, a shared IT vision as well as the communication habits between CIO and CEO can affect IT decisions [4,5]. In addition, trust and the relationship between CIO and CEO can have an impact on the process of reaching a final result [5,6]. Across different industries, various CIO leadership profiles were identified: the IT orchestrator, IT mechanic, IT advisor and IT laggard [7]. In

[1] Corresponding Author: Johannes Thye M.A., research fellow, Hochschule Osnabrück, Caprivistraße 30A, 49076 Osnabrück, Germany; E-mail: johannes.thye@hs-osnabrueck.de.
[2] We use the term CIO for all persons in a leading IT role irrespectively of their position as board member.

addition, two more types of CIOs were found to exist in hospitals: IT managers and system administrators [8]. Also at CEO level, different leadership styles were observed, which embrace the facilitative, participative, result oriented, visionary and adaptive style [9]. Although these styles correspond with a particular way of reaching decisions, some do not give insight into the specific interaction of the CEO and CIO nor do they make assumptions about different decision-making personalities at CEO level. Furthermore, in some cases the categories are not adapted to the health care industry and do not focus on IT decision making. Thus, the question persists, how do CEOs cope with the situation, that they usually have a broad knowledge of many areas but must decide on a highly specialised topic such as hospital IT. This study, therefore, aimed at a) exploring different types of IT decision makers at CEO level, b) identifying hypotheses if trust existed between these different types of CEOs and their CIOs and c) investigating what potential consequences this has on taking risks and innovation capabilities.

2. Methods

In order to answer the research questions, a qualitative study design was chosen since it provides an in-depth understanding of underlying values, mechanism and their complexity. Therefore, semi-structured interviews with hospital CEOs in Germany were conducted. [1] Further reasons for a qualitative approach were: a) to explore the communication patterns instead of confirming existing knowledge, b) to involve the leaders directly and personally, which is difficult to be achieved in quantitative questionnaires, and c) to stimulate further research by developing hypotheses. This strategy is in line with the current methodological literature [10,14]. Experts can provide their specialist knowledge that researchers do not know. Especially semi-structured interviews offer the possibility for the interviewee to come up or consider topics that researchers might not have been taken into account. Therefore, each investigation offers the chance to formulate new questions, theories and suggestions [11]. Also access barriers of high-ranking experts, such as CEOs, like time constraints, a general lack of willingness to provide information or the delegation to personal assistants (who complete the questionnaire) [12], can be overcome and thus an exchange on a personal level be established. Above all these reasons that speak in favour of qualitative methods, they offer the opportunity to get in touch with a group of persons who are hard to reach and to obtain "information between the lines". In this way, unfiltered information can be collected and freedom to explore specific and possibly unexpected special knowledge can be brought off [12].

Although the literature on our research questions is rather scarce we tried to build on existing knowledge as much as possible. The interview guideline was therefore developed on findings from a literature search about decision making and communication in general. Common databases such as ACM, PubMed, SpringerLink, PSYCINFO were researched. Further literature was identified in a snow ball search. An additional google search was conducted. Keywords like CEO, CIO, CIO-CEO relationship, decision making, health IT and synonyms were used and combined to find relevant studies. The studies found focused on reporting structures, IT decision and roles

[1] Further interview results were published in [3]. The methods description of this contribution focuses therefore only on the relevant issues for answering the research questions of this part of the study.

Table 1. Hospital characteristics

ownership	hospital size	system affiliation (hospital in a group)	teaching status
public [n=11]	up to 299 beds [n=3]	yes [n=7]	university hospital [n=1]
private [n=3]	300 to 599 beds [n=7]	no [n=7]	teaching hospital [n=10]
	more than 600 beds [n=4]		no teaching hospital [n=3]

[4,7,9,15-17], IT governance and strategy [4,7,9,18,19] as well as on CIO relationship [4,5,7,20]. Decision making regarding health IT was only poorly covered.

The following topics were addressed by the interviewer in eight open questions: who decides, relationship and collaboration with the CIO, importance of and attitude towards IT, satisfaction with IT development. The guideline was flexible regarding the inclusion of new aspects or evolving questions. Finally, the interviewees were asked to provide background information about themselves and their institution.

In order to obtain a variety of participants, demographics data were used to select different hospital CEOs (Table 1).

Due to the limited availability of hospital CEOs, this study was based on a convenience sample. Fourteen interviews were conducted during the period from 30th May to 11th October 2016 with CEOs from German hospitals. Twelve CEOs were interviewed face-to-face and two via telephone. The interviews lasted between 30 and 75 minutes. Table 2 shows the participant characteristics.

Data collection was finished at the point of saturation, i.e. when the interview results became redundant [12]. The interviews were recorded and then transcribed with MAXQDA 12. MAXQDA is a software with focus on the analysis of qualitative data and text in general and supports the scientific / content wise evaluation of interviews, texts and media. MAXQDA can be used for qualitative, quantitative and mixed method research. The software enabled the interviews to be transcribed, coded and evaluated in a uniform way. Data analysis was initially conducted using categories from the literature in a deductive manner but was extended during the coding process when new categories were added inductively based on the answers. Examples of categories were decision levels, key indicators for decision making or communication and cooperation. An example of a new category was IT projects, because CEOs explicitly reported about single IT projects. Another one was health information systems, which was included in the interviews after the sixth interview.

Table 2. Participant characteristics

gender	age	background	position
female [n=2]	between 32 years and 56	business [n=14]	managing director [n=9]
male [n=12]	years [mean 46, SD 7.26]	medicine [n=1]	business director [n=4]
		nursing [n=4]	procurator [n=1]

3. Results

Pursuant to the information obtained in the interviews about

a) the CEOs' IT knowledge and training,
b) their view on the importance of IT and
c) their attitudes towards IT,

three major types of personalities were found among those who were in a position to make real decisions.

IT savvy CEOs [n=3]: They had received a formal IT education or had been working in an IT department of a hospital. For example, the CEO previously managed IT projects or work as a kind of internal IT consultant. “I accompanied the IT processes at group level before I started here” as CEO [interview 8]. ***IT enthusiastic CEOs*** [n=6]: IT enthusiastic CEOs had no formal IT training, but regarded themselves as very open towards and interested in IT and IT staff / CIO. This type of CEO tried to make themselves acquainted with IT and to catch up on new developments. However, they often did not have enough time for doing so. “[...] throughout the entire process of admission, diagnosis, therapy and discharge [...] IT has a very important supporting role [...] without IT, nothing works or at least not much" [interview 14]. ***IT indifferent CEOs*** [n=3]: Some representatives of this type of CEO had a certain degree of IT knowledge but less knowledge about health IT itself. Health IT played a subordinate role according to their opinion because other topics such as restructuring, corporate finance were regarded as more important or because IT was considered as a mere expense factor. An electronic health record “[…] has been projected for years, but due to a shortage of funds it has never been implemented” [interview 7].

There were another two persons who could not be assigned to any of the three categories because they had no mandate to make global IT decisions and because the strategy was already decided upon. Thus, their IT knowledge and interest was of no importance for any IT decisions.

Based on the CEOs’ statements about the CEO-CIO relationship and collaboration, trust turned out to be a crucial element as represented by the following statements “[...] IT is, to a large extent, also a matter of trust in the CIO, because they can tell me anything” [interview 6] and “you need tremendous trust [...] to the IT service provider, no matter if it is an internal or external provider. That's essential. Decisions will build on trust” [interview 5]. This meant that trust in the CIO was required in any case and particularly when the CEO had no or little IT knowledge. Another factor that emerged from the statements was the role of the language and terminology used. “[...] it is also due to the language. The CIO sometimes uses a bit too much IT jargon [...]” [interview 4]. Not using the same language could therefore become a barrier between the CEO and the CIO. "Each professional group and department has its own specific language and of course also the IT department. In order for us to be able to understand each other, we should hire quite a lot of interpreters and since we do not want to afford them, we are trying to speak one language [...] this is a constant process, so that they do not hide themselves behind the technical terminology as a shield” [interview 13].

These two elements, trust and language, were also mirrored by the three types of CEO personalities. ***IT savvy CEOs*** trusted the information given by their CIOs. The CEOs reported that both spoke the same language. Also, ***IT enthusiastic CEOs*** confirmed to strongly trust the CIO. These statements expressed an even greater degree of trust than the ones of IT savvy CEOs. The IT enthusiastic CEOs also admitted that trust was necessary because they had less IT knowledge. They also mentioned that the language had to be adjusted, however, they would usually understand each other. ***IT indifferent CEOs*** implied that trust in the CIO was less often present. Figure 1 shows the combination of IT knowledge identified and the degree of trust determined for the different types of CEOs. The left diagram in Figure 1 represents the results with regard to the existing trust, the right diagram relates to the trust that would be appropriate due to the level of domain IT knowledge and due to differences in the terminology used.

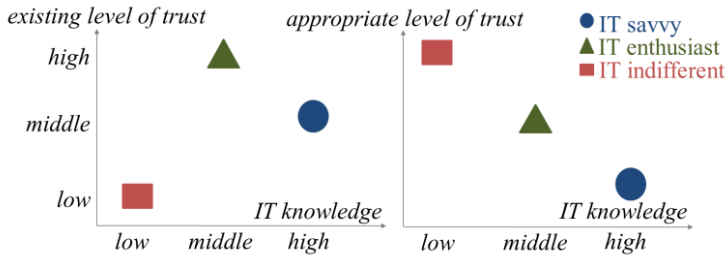


Figure 1. Portfolio matrix of CEO decision makers

IT indifferent CEOs, who had a low level of trust but would need trust in order to make sound IT decisions, were less willing to take a risk. Alike, their innovation capability could be rated lower than that of the other types of CEOs.

4. Discussion

Based on the notion that CEOs are usually the IT decision makers [3], this qualitative study aimed at identifying different types of CEOs. The interviews conducted revealed three major personalities of CEOs: IT savvy CEOs, IT enthusiastic CEOs, and IT indifferent CEOs. In contrast to previous classifications of general leadership styles [9], these types relate to IT, an area of highly specialised knowledge that does not necessarily belong to the competencies of executives. The interviews furthermore underpin the need of trust between CIOs and CEOs as a basis for reliable decisions. The results confirm the finding that CEOs often do not have the necessary IT knowledge, but want someone they can trust in [6,20], which also holds true outside healthcare [5]. A trusted relationship between CEO and CIO can help to place IT topics into the agenda of the top management team [6]. The results also show that CEOs who needed trust in the CIO actually did not have it and vice versa. This circumstance can lead to a vicious circle for IT indifferent CEOs (Figure 2): no IT knowledge is associated with less trust in IT and the CIO, which leads to a rather low willingness to take a risk and less innovative behaviour. As less decisions in favour of IT investments are made, no experience with IT implementations and usage can be gathered. This hypothetic vicious circle needs further research to be corroborated.

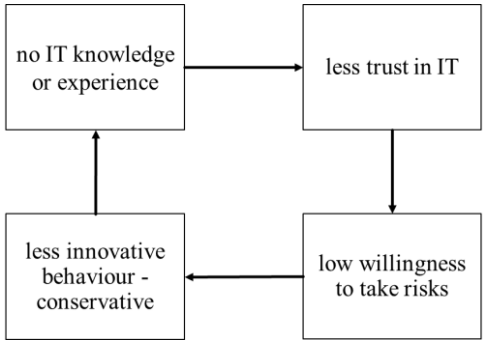


Figure 2: Vicious circle of IT decision making

In order to break of this circle, there seems to be a need to confide in each other [5]. Furthermore, IT knowledge or at least the use of a common language can help to establish a shared understanding [5,20] and to facilitate building trust. Thus, the question arises what IT knowledge, and more generally what, IT core competencies do CEOs need to have and how can they acquire them. Again, this leads to building hypotheses for future work as anticipated and intended by this qualitative study. We thus propose the following hypotheses to be investigated:

1) Little knowledge about health IT and low levels of trust in the CIO / IT lead to a self-reinforcing mechanism that keeps institutions from implementing innovative IT (vicious circle hypothesis).
2) This hypothetic vicious circle can be interrupted by either increasing the IT competence of CEOs (competence hypothesis) or
3) via a common language or terminology (language hypothesis).
4) In both cases a trusting relationship between CEO and CIO would result.

We thus can conclude: This study pointed out three potential types of CEOs based on their IT knowledge, experience and attitude towards IT. It also helped identifying a set of hypotheses which demonstrate the importance of investigating issues of trust, IT competencies and language as mediating factors in taking risks and making decisions.

Fourteen study participants are a rather small group, which certainly limits the generalisation of the findings in terms of quantitative research. A further limitation is the focus on hospitals from western and northern Germany. We addressed this problem by interviewing nevertheless at least one CEO from southern and eastern Germany. Political correctness of answers cannot be eliminated, therefore the personal appraisal of the interviewer was noted. Therefore, caution needs to be exercised by interpreting the results. Nevertheless, the international literature partly supports these findings, which indicates that the hypotheses are also valid beyond Germany and potentially also in the healthcare sector. This is underlined by the following two statements: “CEOs without a technical background often don't know much about IT and therefore want a reliable IT manager they can trust” [5]. “[...] Trust was a key factor underlying the success of personal appeal behaviours. If the CIO had a good track record with [information system] projects and had established a relationship of trust with a peer, then it was likely that the peer would be swayed by personal appeal behaviour [6].”

Still, more research is needed to corroborate the assumptions and test the hypotheses in quantitative studies. Further research should also reflect the view of hospital CIOs to get the entire picture of the relationship and provide insights into reliable structures and processes of sound of IT decision making. Beyond stimulating new research, this study also raised the pragmatic question of how much and what specific IT knowledge do CEOs need to have to decide about complex IT matters.

Acknowledgement

We wish to thank all CEOs for participating in this study. This study was funded by the Ministry of Science and Culture of the State of Lower Saxony (grant: ZN 3062).

References

[1] G. Kuperman, Reflections on AMIA—looking to the future, Journal of the American Medical Informatics Association 20 (2013), e367.
[2] A. Simon, Die betriebswirtschaftliche Bewertung der IT-Performance im Krankenhaus am Beispiel eines Benchmarking-Projektes, in: H. Schlegel (Ed.), Steuerung der IT im Klinikmanagement, Vieweg + Teubner Verlag, Wiesbaden, 2010, pp. 73–90.
[3] J. Thye, U. Hübner, J. Hüsers, B. Babitsch, IT Decision Making in German Hospitals – Do CEOs Open the Black Box? Studies in Health Technology and Informatics 243 (2017), 112–116.
[4] V. Krotov, Bridging the CIO-CEO gap: It takes two to tango, Business Horizons 58 (2015), 275–283.
[5] A. Hütter, T. Arnitz, R. Riedl, Die CIO/CEO-Partnerschaft als Schlüssel zum IT-Erfolg, HMD Praxis der Wirtschaftsinformatik 50 (2013), 103–111.
[6] H.G. Enns, S.L. Huff, C.A. Higgins, CIO Lateral Influence Behaviors: Gaining Peers' Commitment to Strategic Information Systems, MIS Quarterly 27 (2003), 155–176.
[7] D.S. Preston, D.E. Leidner, D.Q. Chen, CIO Leadership Profiles: Implications of CIO Leadership Profiles: Implications of Matching CIO Authority and Leadership Capability on IT Impact, MIS Quarterly Executive 7 (2008), 71–83.
[8] F. Köbler, J. Fähling, H. Krcmar, J.M. Leimeister, IT Governance and Types of IT Decision Makers in German Hospitals, Business & Information Systems Engineering 2 (2010), 359–370.
[9] D. Reeleder, V. Goel, P.A. Singer, D.K. Martin, Leadership and priority setting: the perspective of hospital CEOs, Health policy (Amsterdam, Netherlands) 79 (2006), 24–34.
[10] R.S. Weiss, Learning from Strangers. The Art and Method of Qualitative Interview Studies., The Free Press, New York, 1995.
[11] J. Gläser, G. Laudel, Experteninterviews und qualitative Inhaltsanalyse, VS Verlag für Sozialwissenschaften, Wiesbaden, 2010.
[12] S. Lamnek, C. Krell, Qualitative Sozialforschung, Beltz Verlag, Weinheim, 2016.
[13] N. King, C. Horrocks, Interviews in Qualitative Research, SAGE Publications Ltd, London, 2010.
[14] J.W. Creswell, Qualitative inquiry and research design: Choosing among five approaches, SAGE, Los Angeles, 2013.
[15] R.B. Banker, N. Hu, P.A. Pavlou, J. Luftman, CIO Reporting Structure, Strategic Positioning and Firm Performance, MIS Quarterly 35 (2011), 487–504.
[16] H. Moghaddasi, A. Sheikhtaheri, CEO is a vision of the future role and position of CIO in healthcare organizations, Journal of medical systems 34 (2010), 1121–1128.
[17] F. Köbler, J. Fähling, J.M. Leimeister, H. Krcmar, How German Hospitals Govern IT - An Empirical Study, Proceedings of the 17th European Conference on Information Systems (ECIS); Verona, Italy (2009), 317.
[18] P. Weill, J.W. Ross, IT Governance, Harvard Business School Press, Boston, 2004.
[19] A.M. Johnson, A.L. Lederer, IS Strategy and IS Contribution: CEO and CIO Perspectives, Information Systems Management 30 (2013), 306–318.
[20] D.F. Feeny, B.R. Edwards, K.M. Simpson, Understanding the CEO/CIO Relationship, MIS Quarterly 16 (1992), 435–448.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-47

Improving Fluid Management in Critical Care – Towards the ICU of the Future

Katharina BERGMOSER[a,b,1], Lucas PFLANZL-KNIZACEK[a,c], Matthias HAFNER[a,d], Gernot SCHILCHER[e], Christian BAUMGARTNER[b]
[a]*CBmed – Center for Biomarker Research in Medicine, Graz, Austria*
[b]*Graz University of Technology – Institute of Health Care Engineering with European Testing Center of Medical Devices, Graz, Austria*
[c]*Medical University of Graz, Division of Endocrinology and Diabetology, Austria*
[d]*Medical University of Graz – Department of Anesthesiology and Intensive Care, Austria*
[e]*Medical University of Graz – Intensive Care Unit, Department of Internal Medicine, Austria*

Abstract. Background: The calculation of daily fluid balances is essential in perioperative and postoperative fluid management in order to prevent severe hypervolemia or hypovolemia in critically ill patients. In this context, modern health information technology has the potential to reduce the workload for health care professionals by not only automating data collection but also providing appropriate decision support. Objectives: Within this work, current problems and barriers regarding fluid balancing in cardiac intensive care patients are outlined and improvement activities are specified. Methods: Literature research and qualitative interviews with health care professionals were conducted to assess the state-of-the-art technological setting within an intensive care unit. Results: An example case shows that interconnecting not only devices but also wards can facilitate daily clinical tasks. Conclusion: Smart devices and decision support systems can improve fluid management. Several technologies, which today are sometimes still considered to be futuristic, are in fact not that far away or already available. However, they need proper implementation with respect to intensivists', nurses' and patients' needs.

Keywords. Intensive Care, Information Technology, Clinical Decision Support

1. Introduction

As in other areas, modern technologies find more and more their way into healthcare and do not stop in front of hospitals. At the moment, also intensive care units (ICU) are on move to interconnected institutions, where the Internet of Things (IoT) plays an important role in meeting the steadily rising trend for personalization. However, despite all advantages of digitalization, also new challenges due to e.g. novel measurement methods or big data arise when modernizing an existing ICU by introducing smart devices and innovative sensors for a nearly continuous monitoring of vital parameters or biomarkers [1].

An area, where especially interconnection of devices could decrease the workload for health care professionals (HCP) and increased data quality could enhance treatment individualization, is intra- and postoperative fluid management. The administration of

[1] Corresponding Author: Katharina Bergmoser, CBmed – Center for Biomarker Research in Medicine, Stiftingtalstrasse 5, 8010 Graz, Austria, E-Mail: katharina.bergmoser@cbmed.at

fluids in critically ill patients is widely discussed in terms of type, amount, timing and outcomes [2-7]. However, recent works indicate that in cardiac intensive care the amount of fluids rather than the type needs to be chosen with care and that the fluid balance – as a strong indicator for the current fluid therapy's efficacy – should closely be monitored [8]. The fluid balance is defined as the difference between all fluid intakes and all ongoing fluid losses [9]. Ideally, intakes and losses balance each other and the overall fluid balance is zero. In critically ill patients, fluid dysbalances should be avoided [10] and the presence of fluid overload might be an important clinical marker when estimating the risk for the development of an acute kidney injury (AKI) [11]. Within ICUs, setting targets for daily fluid balancing plays an important role in guiding removal of excess fluids [12]. In case of a severe fluid overload, under certain circumstances the use of renal replacement therapy may be preferable to the administration of diuretics [12].

For determining the fluid balance, administered fluids and ongoing losses have to be recorded accurately. A study by Bashir et al. [13] showed that in addition to the recorded fluids up to 1.5 liters of "hidden" fluids are administered to a patient each day. In contrast to healthy adults, patients might have additional sources of water and electrolyte losses resulting from e.g. bleedings or fever. In daily fluid balance calculation at the ICU, at least an insensible fluid loss of 10ml/kg/day should be considered in non-intubated patients [14]. During surgery, urinary fluid losses and losses through perspiration sum up to approximately 0.5-1.0 ml/kg/h [15].

The objective of this work is to outline current problems and barriers in fluid management, especially in fluid balancing of cardiac intensive care patients from a technological point of view. Furthermore, an improved setting within a cardiac ICU is sketched, which facilitates fluid management in critical care.

2. Methods

2.1. Knowledge base for developing an integrative ICU concept

Qualitative interviews with HCP (n=12) differing in years of experience were conducted in order to obtain detailed information on the typical settings and workflows in critical care. Discussions with nurses and physicians working in intensive care settings were initiated in order to identify current obstacles and generate ideas for improvement. A semi-structured literature research (keywords e.g. fluid management, fluid balance, intensive care, health information technology, decision support, automation, state of the art, trend, future) revealed how state-of-the-art as well as future technologies could contribute to a smart and more user-friendly environment.

2.2. Fluid management: Patient-specific parameters in a current intensive care setting

Patients staying at ICUs are typically surrounded by a range of medical devices, which are necessary for either monitoring the health status or delivering therapies. Figure 1 shows several devices, whose parameters and measurements need to be considered in fluid balancing. Depending on the type and severity of the disease, a large number of parameters are assessed either frequently or if necessary. Recorded data contain numerical values such as the amount of urinary excretion or ordinal-scaled values for defining severity levels and disease stages. An example for routinely used scales are the RIFLE and AKIN criteria for staging AKI [16].

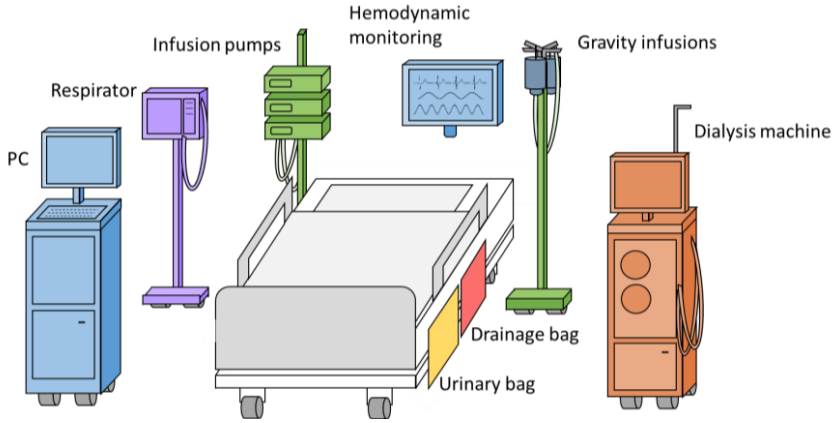


Figure 1. Devices important for fluid balancing stationed around a critical care patient's bed.

Hemodynamic vital parameters such as heart rate, blood pressure or ventilation as well as administered fluids and medications are recorded automatically and transmitted to a patient data management system (PDMS). Fluid losses are still entered manually using fill level readings on fluid containers. In general, the manual calculation of fluid balances is time-consuming but accurate documentation seems to be helpful in AKI prevention [17]. In catheterized patients, urinary losses are usually recorded several times a day, whereby an hourly monitoring is probably associated with less development of fluid overload and a better detection of AKI [18]. Fluid losses such as sweating or insensible losses via skin and lungs are often not taken into account. However, in case of high fever an additional fluid loss of approximately 500 ml/day can be assumed [14].

To our knowledge, the preoperative fluid balance of patients undergoing elective cardiac surgery is not assessed but might have an impact on intraoperative fluid management. The perioperative fluid balance is often not available in the ICU caregivers' PDMS due to a missing interface to the intraoperative documentation system. This leads to assuming neutral balances for patients arriving at ICU. Therefore, the initiated postoperative fluid therapy is not adequate because of a – sometimes – strongly under- or overestimated postsurgical fluid status [15]. In general, there seems to be a need for improving the transmission of intraoperative fluid management data from the operating room to the ICU at patient handover in terms of information quality and availability [19].

3. Results

The basis for an efficient fluid management in intensive care are availability, correctness, completeness and user-friendly display of relevant data. In order to achieve that, the following aspects should be considered when moving towards the next-generation ICU.

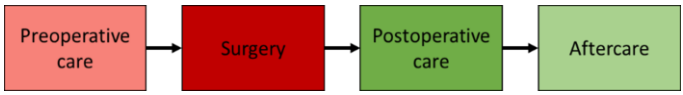


Figure 2. An elective cardiac patient's way through the hospital.

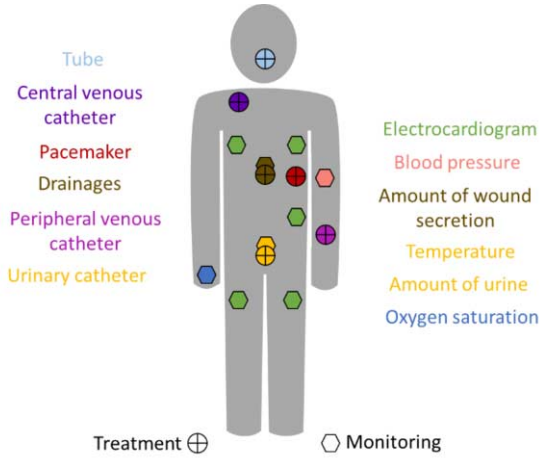


Figure 3. Common accesses and data capture sites in cardiac intensive care patients.

3.1. Closing interface gaps and establishing sensor networks

During an elective cardiac surgery patient's stay at the hospital, he or she passes several units (see Figure 2). It should be ensured, that all relevant information regarding the patient's stay at the current ward is available to the following one at patient handover at the latest.

Especially assistive or not routinely used devices around the patient bed such as extracorporeal membrane oxygenation (ECMO) or dialysis machines are often not connected to the PDMS and work as standalone systems where manual data transfer is needed. It is important to close these interface gaps to reduce workload for caregivers and to avoid errors due to manual data entry. Furthermore, the more data from different sources are integrated, e.g. preoperative baseline parameters or results from blood gas analyses, the better the basis for decision support systems (DSS) and algorithm-driven care [20] is. For instance, real-time data from continuous renal replacement therapy (CRRT) with interfaces to electronic health records (EHR) represent an important feedback loop for estimating the amount of medication filtered by the dialysis machines and could be used for an (semi-)automated adjustment of infusion rates [21].

Intensive care patients are usually equipped with a range of monitoring and treatment accesses, depending on the disease and the standard operating procedures of the respective hospital. Some of them are displayed in Figure 3.

Cable-bound sensors for monitoring purposes should be replaced by wireless solutions whenever their application makes sense and contributes to a more user-friendly environment [22]. Ad-hoc sensor networks should be established with respect to safety and security [23]. Figure 4 shows how intelligent sensor networks can build the basis for an individualized fluid management.

3.2. Introducing decision support systems but avoiding alarm fatigue and social silos

In critical care, HCP are faced with a range of measurements, displayed in form of either a current measurement value or time series. Simultaneous interpretation of up to 10 vital parameters is often necessary for assessing a patient's current condition [24]. Therefore,

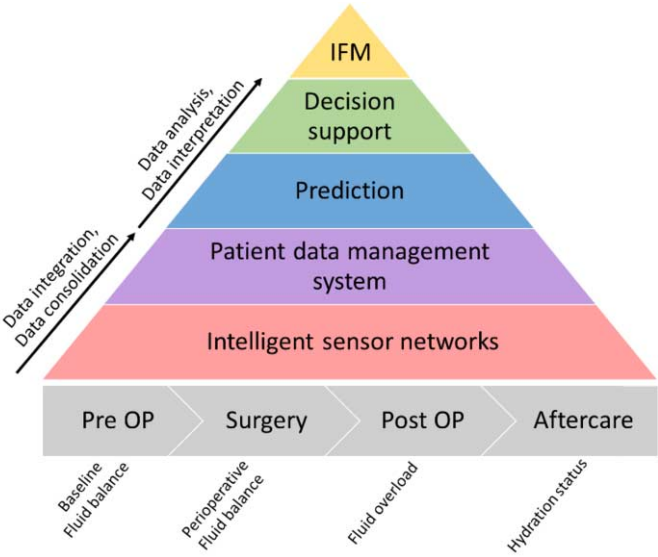


Figure 4. Intelligent sensor networks can build the basis for individualized fluid management (IFM) by enabling cross-departmental and user-friendly collection of vital parameter data.

it is important to reduce the amount of displayed parameters to the currently important ones. Clinical DSS based on machine learning algorithms might aid in selecting them and ensuring an intuitive and meaningful data representation. Additional parameters processed by background algorithms might be used for indicating changing health conditions and suggesting treatment options.

It is beyond debate, that acoustic alarm in case of worsening vital parameters is important. However, an appropriate and not too sensitive adjustment of the limit values for alarms leads to a reduction in physiologic monitor alarms [25] and therefore contributes to avoidance of alarm fatigue and facilitation of patient recovery [26-28]. Furthermore, alarm sounds should be melodic and easy to learn [29].

Comprehensive DSS are designed with respect to the actual user group, integrating all available data, providing prediction and giving assistance in planning next treatment steps. Nevertheless, cross-departmental communication among HCPs is essential and should be fostered to maintain treatment quality and avoid social silos [30].

3.3. The integrative ICU – an exemplary use case

John is 65 years old and has to undergo elective coronary artery bypass grafting (CABG). At the day of admission, besides several intake interviews with HCP and a baseline blood count, also a baseline fluid balance of -500 ml is determined using bioelectrical impedance analysis [31]. The obtained parameters are automatically transmitted to a central PDMS system. At the cardiac surgery unit, the integrated DSS suggests options for how to best reach a neutral fluid balance until surgery – which takes place the next morning – considering dinner, preoperative fasting and estimation of losses to be expected during night.

At surgery initiation, an algorithm estimates John's fluid balance based on the administered fluids since admission, allowing the anesthetist to keep track of his actual

fluid status. During surgery, important blood electrolytes such as sodium, potassium or chloride are permanently monitored using a wireless inline-measurement system reducing blood loss. Those measurements as well as the administered fluids and recorded losses are automatically stored in the surgery protocol within the PDMS. Fluid losses and gains through the heart-lung machine are taken into account as well.

After surgery, John is transferred to the cardiac ICU, where a bed has already been assigned to him automatically by the bed management system being part of the PDMS. Each bed is equipped with a tablet, which provides a clear overview of a selection of John's vital parameters. Besides the measurements from the inline system and hemodynamic parameters, an estimation of the postoperative fluid balance is displayed. John's postoperative values are in range, accept for a slight fluid overload of +700 ml. The flow rates of administered fluids and medications as well as those within drainages and the urinary bag are continuously measured and automatically updated in the PDMS.

In the evening, the urinary flow rate steadily decreases and the inline-measurement of potassium shows upward tendency. The DSS recommends the measurement of kidney-specific laboratory parameters and suggests first countermeasures. Since the parameters are not yet in a critical range, the responsible HCP are notified via smartphone and smartwatch messages. Compared to baseline measurements, John's vital parameters worsen during the next day. At the evening of the second postoperative day, his fluid balance increased to +4000 ml and the prediction algorithm shows a further upward trend. Since the intravascular volume is in range, the DSS indicates an increased risk for AKI and suggests the consultation of a nephrologist, which can directly be video-called or messaged using John's bedside tablet. Later in the course of his stay, a CRRT has to be started whereby the parameters of the dialysis machine are set using the values suggested by the DSS. All status variables of the dialysis machine are permanently monitored and wirelessly sent to the PDMS.

In the following days John recovers, the CRRT is terminated since diuresis and blood parameters normalized and the inline-electrolyte measurement system is removed. With a slightly positive balance of +200 ml, John is handed over to aftercare, where HCP already received a digital summary of John's stay.

4. Discussion

Mobile devices and apps are important tools not only for documenting and monitoring a patient's health status but also for consulting guidelines or obtaining decision support [32]. During perioperative care, DSS are already used for estimating e.g. complication risks or severity of illness [33]. In fluid management of the near future, introducing feedback loops using hemodynamic parameters will play an important role in prevention of hypo- and hypervolemia [34]. First approaches for guiding fluid therapy have already been published [35,36].

Some of the technologies described in the exemplary patient case are already available. For instance, the inline-measurement of blood components – which has the advantages of no blood loss and real-time data – is possible [37]. A sensor network dedicated to the use within critical care has already been developed by Shnayder et al. [23]. However, in the far future those networks might partly be replaced by all-in-one solutions [38]. Sensors themselves will reduce in size and be available in form of wearables [39]. Medical devices within the ICU might even be replaced by ingestible smart pills [40].

However, in fluid management integrative and patient-oriented systems based on our proposed concept will gain in importance. A first essential step in tackling current implementation barriers is the introduction of license-free and state-of-the-art interfaces enabling devices to transmit data to the PDMS. User-friendly sensing, automated and secure data capture, the introduction of intelligent and non-unobtrusive DSS but also the maintenance of face-to-face discussions and collaboration will be necessary to ensure an optimal monitoring of patients on their journey through the hospital.

Acknowledgements

"CBmed" is cooperating together with B.Braun Melsungen AG for this research project. Work done in "CBmed" was funded by the Austrian Federal Government within the COMET K1 Centre Program, Land Steiermark and Land Wien.

References

[1] V. Tresp, J.M. Overhage, M. Bundschus, S. Rabizadeh, P.A. Fasching, Y. Shipeng, Going digital: a survey on digitalization and large-scale data analytics in healthcare, Proceedings of the IEEE 104(11) (2016), 2180-2206.
[2] J. Benes, M. Kirov, V. Kuzkov, M. Lainscak, Z. Molnar, G. Voga, X. Monnet, Fluid therapy: double-edged sword during critical care?, BioMed Research International 2015 (2015).
[3] V.A. Bennett, M. Cecconi, Perioperative fluid management: from physiology to improving clinical outcomes, Indian Journal of Anaesthesia 61(8) (2017), 614-621.
[4] V. Chatrath, R. Khetarpal, J. Ahuja, Fluid management in patients with trauma: restrictive versus liberal approach, Journal of Anaesthesiology Clinical Pharmacology 31(3) (2015), 308-316.
[5] L.H. Navarro, J.A. Bloomstone, J.O. Auler Jr, M. Cannesson, G.D. Rocca, T.J. Gan, M. Kinsky, S. Magder, T.E. Miller, M. Mythen, A. Perel, D.A. Reuter, M.R. Pinsky, G.C. Kramer, Perioperative fluid therapy: a statement from the international fluid optimization group, Perioperative Medicine 4(3) (2015).
[6] D.A. Reuter, D. Chappell, A. Perel, The dark sides of fluid administration in the critically ill patient, Intensive Care Medicine 2017 (2017).
[7] G. Minto, M.G. Mythen, Perioperative fluid management: science, art or random chaos?, British Journal of Anaesthesia 114(5) (2015), 717-721.
[8] E. Bignami, M. Guarnieri, M. Gemma, Fluid management in cardiac surgery patients: pitfalls, challenges and solutions, Minerva Anestesiologica 83(6) (2017), 638-651.
[9] D.N. Lobo, A.J. Lewington, S.P. Allison, Basic concepts of fluid and electrolyte therapy, Bibliomed-Medizinische Verlagsgesellschaft mbH, Melsungen, 2013.
[10] V. Balakumar, R. Murugan, F.E. Sileanu, P. Palevsky, G. Clermont, J.A. Kellum, Both positive and negative fluid balance may be associated with reduced long-term survival in the critically ill, Critical Care Medicine 45(8) (2017), e749-e757.
[11] N. Salahuddin, M. Sammani, A. Hamdan, M. Joseph, Y. Al-Nemary, R. Alquaiz, R. Dahli, K. Maghrabi, Fluid overload is an independent risk factor for acute kidney injury in critically ill patients: results of a cohort study, BMC Nephrology 18(1) (2017), 45.
[12] M.E. O'Connor, S.L.Jones, N.J. Glassford, R. Bellomo, J.R. Prowle, Defining fluid removal in the intensive care unit: a national and international survey of critical care practice, Journal of the Intensive Care Society 18(4) (2017), 282-288.
[13] M.U. Bashir, A.Tawil, V.R. Mani, U. Farooq, M.A. DeVita, Hidden obligatory fluid intake in critical care patients, Journal of Intensive Care Medicine 32(3) (2017), 223-227.
[14] P. Cox, Insensible water loss and its assessment in adult patients: a review, Acta Anaesthesiologica Scandinavica 31(8) (1987), 771-776.
[15] M. Rehm, N. Hulde, T. Kammerer, A.S. Meidert, K. Hofmann-Kiefer, State of the art in fluid and volume therapy: a user-friendly staged concept. English version., Der Anaesthesist 2017 (2017).
[16] KDIGO Work Group, KDIGO clinical practice guideline for acute kidney injury, Kidney International Supplements 2(1) (2012).

[17] A. Davies, S. Srivastava, W. Seligman, L. Motuel, V. Deogan, S. Ahmed, N. Howells, Prevention of acute kidney injury through accurate fluid balance monitoring, BMJ Open Quality 6(2) (2017).
[18] K. Jin, R. Murugan, F.E. Sileanu, E. Foldes, P. Priyanka, G. Clermont, J.A. Kellum, Intensive monitoring of urine output is associated with increased detection of acute kidney injury and improved outcomes, Chest 152(5) (2017), 972-979.
[19] N. Segall, A.S. Bonifacio, R.A. Schroeder, A. Barbeito, D. Rogers, D.K. Thornlow, J.Emery, S. Kellum, M.C. Wright, J.B. Mark, Can we make postoperative patient handovers safer? A systematic review of the literature, Anesthesia & Analgesia 115(1) (2012), 102-115.
[20] T.J. Pollard, L. A. Celi, Enabling machine learning in critical care, ICU Management & Practice 17(3) (2017 Fall), 198-199.
[21] W.R. Clark, M. Neri, F. Garzotto, Z. Ricci, S.L. Goldstein, X. Ding, J. Xu, C. Ronco, The future of critical care: renal support in 2027, Critical Care 21(92) (2017).
[22] M. Paksuniemi, H. Sorvoja, E. Alasaarela, R. Myllylä, Wireless sensor and data transmission needs and technologies for patient monitoring in the operating room and intensive care unit, Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference 5 (2005), 5182-5185.
[23] V. Shnayder, B. Chen, K. Lorincz, T. Fulford-Jones, M. Welsh, Sensor networks for medical care, Harvard Computer Science Group Technical Report TR-08-05, http://nrs.harvard.edu/urn-3:HUL.InstRepos:24829604, last access: 29.01.2018.
[24] F.A. Drews, Patient monitors in critical care: lessons for improvement, Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 3: Performance and Tools) (2008).
[25] K.C. Graham, M. Cvach, Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms, American Journal of Critical Care 19(1) (2010), 28-37.
[26] L. Johansson, I. Bergbom, B. Lindahl, Meanings of being critically ill in a sound-intensive ICU patient room – a phenomenological hermeneutical study, The Open Nursing Journal 6 (2012), 108-116.
[27] J.A. Alasad, N.A. Tabar, M.M. Ahmad, Patient's experience of being in intensive care units, Journal of Critical Care 30(4) (2015), 859.e7-859.e11.
[28] M.A. Pisani, R.S. Friese, B.K. Gehlbach, R.J. Schwab, G.L. Weinhouse, S.F. Jones, Sleep in the intensive care unit, American Journal of Respiratory and Critical Care Medicine 191(7) (2015), 731-738.
[29] A.N. Wee, P.M. Sanderson, Are melodic medical equipment alarms easily learned?, Anesthesia and Analgesia 106(2) (2008), 501-508.
[30] M. Leslie, E. Paradis, M.A. Gropper, S. Kitto, S. Reeves, P. Pronovost, An ethnographic study of health information technology use in three intensive care units, Health Services Research 52(4) (2017), 1330-1348.
[31] M.Y. Jaffrin, H. Morel, Body fluid volumes measurements by impedance: a review of bioimpedance spectroscopy (BIS) and bioimpedance analysis (BIA) methods, Medical Engineering & Physics 30(10) (2008), 1257-1269.
[32] C.L. Ventola, Mobile devices and apps for health care professionals: uses and benefits, Pharmacy and Therapeutics 39(5) (2014), 356-364.
[33] A. Belard, T. Buchman, J. Forsberg, B. K. Potter, C. J. Dente, A. Kirk, E. Elster, Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care, Journal of Clinical Monitoring and Computing 31(2) (2017), 261-271.
[34] A. Joosten, B. Alexander, A. Delaporte, M. Lilot, J. Rinehart, M. Cannesson, Perioperative goal directed therapy using automated closed-loop fluid management: the future?, Anaesthesiology Intensive Therapy 47(5), 517-523.
[35] K. Bergmoser, L. Pflanzl-Knizacek, S. Langthaler, C. Baumgartner, Describing cardiac ICU patient's fluid transfer characteristics using system analysis – a proof of concept, Medical Fluids 5(1) (2017), 87-88.
[36] J. Rinehart, C. Lee, C. Canales, A. Kong, Z. Kain, M. Cannesson, Closed-loop fluid administration compared to anesthesiologist management for hemodynamic optimization and resuscitation during surgery: an in vivo study, Anesthesia & Analgesia 117(5) (2013), 1119-1129.
[37] J.J. Fox, T.H. Clutton-Brock, Evaluation of a patient-dedicated blood gas analyser, Critical Care 2015, 19(Suppl 1) (2015), P261.
[38] F. Michard, Innovations in monitoring: from smartphones to wearables, ICU Management & Practice 17(3) (2017), 148 -150.
[39] F. Michard, Hemodynamic monitoring in the era of digital health, Annals of Intensive Care 6(15) (2016).
[40] C. Ince, Intensive care medicine in 2050: the ICU in vivo, Intensive Care Medicine 43(11) (2017), 1700-1702.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-55

Towards a Single Data Exchange Standard for Use in Healthcare and in Clinical Research

Jozef AERTS[a,1]
[a]*University of Applied Sciences FH Joanneum, Institute of eHealth, Graz, Austria*

Abstract. Background: The development processes of data exchange standards for use in healthcare are very different from those used in clinical research. Healthcare data standards are traditionally developed by the Health Level 7 (HL7) organization, whereas those for use in clinical research are mostly developed by the Clinical Data Interchange Standards Consortium. No alignment of these standards has so far taken place. Objectives: Due to the increasing use of electronic health records as primary source in clinical research, it becomes necessary to align these standards, not only the semantic standards, but also the data exchange standards (formats) themselves. Methods: Mutual feature gaps between ODM and FHIR are investigated. Results: A transition path how the HL7-FHIR standard and the CDISC-ODM transport standard can grow into a single standard for use both in healthcare and in clinical research is presented.

Keywords. HL7, FHIR, CDISC, ODM, alignment.

1. Introduction

Information standards to be used in healthcare are traditionally developed by the "Health Level 7" (HL7) organization [1]. The first standard for data exchange was the HL7-v2 message standard, which is still very much used, especially for exchange of data within a single hospital or hospital organization. The use of these messages to exchange data between different organizations is however rare. The HL7-v3 standard was developed to overcome some of the limitations of HL7-v2, the most prominent one being the fact that HL7-v2 is only suitable for messages, and not for documents [2]. This was a huge problem, as care providers are used to exchange information with colleagues by means of "letters", i.e. documents. HL7-v3 however became only successful for documents, especially in the form of CDA (Clinical Document Architecture), which is also the basis of the interoperable health records system ELGA in Austria. Although a number of HL-v3 messages have been developed [3], they have never been very successful. This is probably due to the complexity of HL7-v3, which is based on the RIM (Reference Information Model), in which every piece of information need to be modelled either as an "act", a "participation", a "role", an "entity", an "actrelationship" or a "rolelink" or one of the subclasses of these. This made HL7-v3 hard to implement for IT people in software [4]. Therefore, a new standard "FHIR" (Fast Healthcare Interoperability Resources) [5] has been developed by HL7, encompassing both

[1] Corresponding Author: Jozef Aerts, University of Applied Sciences FH Joanneum, Institute of eHealth, Eckertstrasse 30i, 8020 Graz, E-Mail: Jozef.Aerts@fh-joanneum.at.

messages and documents, and using the most modern technologies such as RESTful web services [6] and different technical implementations such as JSON, XML and Turtle [7].

At the other side, CDISC developed the ODM (Operational Data Standard) for exchange of data within clinical research [8]. Its model is still mostly based on the paradigm of collection of data using paper forms (CRF: Case Report Form) during visits, or the electronic version of this (EDC: Electronic Data Capture) at the best. There is only one technical implementation (XML), which is not very suitable yet for exchange of data coming from wearables and other devices, nor for data originating from electronic health records, nor for remote trials in which there are essentially no visits. The latter are however recent evolutions, for which ODM was originally never designed for.

ODM has been very successful in the last 15-20 years, and is being used "end-to-end" [9], except for electronic submissions to the regulatory authorities. These still require data to be submitted in a very old binary format, the "SAS Transport 5" format [10], this although a variation of ODM (Dataset-XML) has exactly been developed for this use case [11].

2. Methods

Although HL7-FHIR and CDISC-ODM are rather different data exchange standards, there is a strong desire in specific parts of the CDISC as well as the HL7 organization to come to a single standard. The reason for that is that it is obvious that in future, large part of the data used in clinical research will either come from electronic health records, or will be shared between healthcare and research [12]. Coming to such a single standard is not easy, as not only technical issues need to be overcome, but also mental issues, such as a mutual understanding about the differences between healthcare and clinical research. For example, clinical research is "protocol driven" [13], meaning that is exactly predefined which tests need to be executed, which questions need to be asked to the patient. In healthcare, the treating physician in many cases acts "event driven", i.e. takes decisions about which tests to be performed, which treatment to be followed on basis on events that occur, such as an observation during a visit, an outcome of an earlier test, etc.

Also, the clinical research world is rather conservative, and the players in the field (pharma companies, Clinical Research Organizations (CROs), EDC vendors) are very reluctant to have any changes in the standards they use or need to use. Whereas the step from paper to EDC, and the use of the ODM standard was already a huge step for many (and which is still not completed), a move to FHIR (or FHIR-like) as a format, even if FHIR would completely be suitable for clinical research, would be another huge step. Thus, a transition path will be necessary in which FHIR and ODM evolve in the same direction and learn from each other to finally become a single standard. The time frame for this can only be estimated, but the author assumes this time frame will be 10 years or more.

HL7-FHIR resources are still in development, with different "STU"s (Standard for Trial Usage) as milestone stages [14]. Every developed FHIR resource has a "maturity level", ranging from 0 to 5 [15]. For example, the "Patient" resource currently has a maturity level 5 (the highest), whereas the "AdverseEvent" resource has the maturity level 0 (the lowest). This is interesting, as "adverse event" is one of the most important concepts in clinical research: the first requirement for a new drug of treatment is that it

is safe, i.e. that the number and kind of adverse events is low or very low and that the benefit / risk ratio is as high as possible.

In this paper, we report on our evaluations of both the FHIR and ODM standard, with a focus on the development of ODM Version 2 (ODMv2), which will in future replace the current ODM standard Version 1.3.2. ODMv2 will use a good number of concepts from FHIR, in an attempt to make a first step into the right direction with the long term goal that FHIR and ODM can become one single standard. We also report on a number of initiatives in the FHIR community to make FHIR more suitable for clinical research, as can be seen from a new category "public health and research" in FHIR STU3.

Table 1: Features supported by FHIR but absent or not supported in ODM

Feature	Description / Comment
Resources	ODM is not based on resources. However, when the content within a "Form" or "ItemGroupDef" logically belongs together, these can be compared to a "resource". For example, the CDASH forms [16] each describe logical grouping of terms belonging together and that can be compared to a "Resource". Example: CDASH form "Adverse Event" with FHIR resource "AdverseEvent". (also see "Profiles")
RESTful web services	Although a number of EDC vendors have developed RESTful web services for exchange of study metadata and data [17,18], unlike in FHIR [19], there is no standardized API.
Documents and Messages	CDISC ODM does not distinguish between messages and documents. A CDISC ODM file can both be used as a message or document. In FHIR, unlike in HL7-v2 and v3, the difference between a document and a message has become very small, both are "bundles" of resources [20] with either the type being "document" or "message". In the latter case, the first resource in the bundle must be "MessageHeader" whereas in the former case it must be "Composition".
Profiles	Specific use cases of FHIR resources can be described in "profiles" [21]. For example, a "vital signs profile" has been developed [22], describing a set of FHIR "Observation" instances, each defined by a LOINC [23] code. CDISC ODM has no such construct, as it has been regarded as out of the scope of ODM. However, CDISC-CDASH [16] defines semi-standardized forms such as a "vital signs form", describing the components of typical vital signs measurements such as systolic and diastolic blood pressure, body height and weight, pulse, etc. Unfortunately, CDASH does not provide LOINC codes for such measurements. The CDASH forms are however also available in ODM format. Furthermore, a number of people within the CDISC organization have developed "Biomedical concepts" (BCs) [24] which surely have a relationship with FHIR resources. For example, for the BC "systolic blood pressure" [25], it not only describes the test itself, but also the body position in which the measurement was taken (standing, sitting, supine), and the usual unit (mmHg). This compares very well to the "structure definition" of the component "systolic blood pressure" LOINC code 8480-6) in the FHIR profile "vital signs" [22]. Remark that currently, CDISC controlled terminology [26] does not allow the use of UCUM notation for units, whereas UCUM [27] is well established in the HL7 world. This kind of differences may become considerable hurdles when trying to come to a single standard.
Lack of semantics	ODM does not describe what needs to be done. It just describes the framework in which things can be done. The only semantics that is described by ODM is "study", "studyevent" (visit/encounter)", "form", "subform", "item/question", "skip condition"," code list" and "calculation method". ODM does not know what a "laboratory test" is, it only provides the framework to define one.
Different technical implementations	CDISC-ODM only has an XML implementation, whereas for HL7-FHIR, the standard can be implemented as XML, JSON or Turtle [7].
Distributed data	HL7-FHIR supports and promotes the use of distributed data. Single resources of a patient do not necessarily need to reside on the same server, but can be located on any FHIR server anywhere in the world. CDISC-ODM does not support this at all.

Table 2. Features supported by CDISC-ODM but absent or not supported in HL7-FHIR

Feature	Description / Comment
Multi-language support	From the start on, ODM contained multi-language support, and this feature was extended at each version update. The idea is that when a study is being developed, every important information point can be made available in any of the languages of the sites where the study will be conducted. This e.g. means that a single question or item like "systolic blood pressure" can also be defined in German ("systolischer Blutdruck") as sibling elements in the ODM. In FHIR, this would require a different instance of the resource "Questionnaire". Remark that in ODM, the identifier of the data point is independent from the language, so that data can be compared between languages. For example for "sex", the questions and enumerated answers may be in the local language, such as "Weiblich" and "Männlich" for the German language, but the captured values will be stored in a language-independent way in the database, such as "F" and "M".
Predefined data types	FHIR resources do not describe which data type for a data point is expected. This is logical as most FHIR resources describe data that was already captured, whereas the study definition part in ODM defines data that need to be captured. For example, a typical CDASH form definition in ODM [16] will define that the data point for "systolic blood pressure" is expected to be an integer. On the other hand, the data type of the captured data point will usually not appear in the ODM clinical data part of ODM, as it was already defined in the study definition part, where both are connected by a unique identifier within the study, the so-called "OID" which has a completely different structure and meaning than OIDs in healthcare [28]. In FHIR however, the data type appears in the resource itself as a variation of "value[x]", e.g. "valueQuantity", "valueString", "valueDatetime". There are 11 such "data types", whereas these are even more granular in ODM which counts 21 data types.
Audit records	CDISC ODM allows to describe data records that are fully 21 CFR Part 11 compliant by the use of "audit records" [9]. Audit records are however not described by FHIR. FHIR has an "AuditEvent" resource, but it has a different meaning. This is also very important for the ODM use case of archival [9].

3. Results

First, we made an inventory of differences in functionality between FHIR and ODM. Table 1 list a number of features that are found in FHIR but are not absent in ODM versions 1.3.2. Table 2 lists a number of features that are present in ODM but are not present or not supported in FHIR (STU3).

So, when trying to come to a single data exchange standard for as well healthcare as clinical research, it is clear that the best way to do so is to add features in ODM that are supported in FHIR, and add features to FHIR that are supported by ODM.

Leroux et al. made comparison and a mapping between CDISC-ODM and HL7-FHIR [29]. One of the statements in this paper is that "ODM is ill-suited for advancing the semantic interoperability solution". This is correct as ODM essentially describes a framework about "things" that are planned and "things" that happened according to that plan. CDISC did develop a large amount of controlled terminology [26], in first instance for electronic submissions to regulatory authorities, but this controlled terminology is mostly incompatible with what is used in healthcare informatics. For example, CDISC developed lists of as well laboratory tests, vital signs tests and even lists of microorganisms, but these are incompatible with LOINC and SNOMED. They are unfortunately also not mappable to these systems, as they are meant for post-coordination whereas most of the controlled terminology in healthcare informatics is pre-coordinated. When looking at the work done so far, especially the work of Leroux et al. [29] is of major importance. A summary of their mapping between FHIR and ODM is depicted in Table 3:

Table 3. Mapping between FHIR and ODM (summary) according to Leroux et al.

HL7-FHIR Resource	CDISC ODM Element	Comments
CarePlan	Study	Both the resources "CarePlan" and "Study" refer to things that are planned, but with different purposes. Plans of care can change during the treatment period, whereas studies are not meant to change during the study duration. In HL7-FHIR, plans of care can be nested, i.e. a care plan can consist of several sub-plans. This feature is not present in CDISC-ODM.
Questionnaire	FormDef / ItemGroupDef / ItemDef	The HL7-FHIR resource "Questionnaire" describes a list of questions to patients. Answers to these questions can be of several types, which match rather well with the data types in CDISC-ODM for "ItemDef". Both HL7-FHIR as well as ODM allow to define groupings of items. In CDISC-ODM however, "FormDef" is not limited to questions to patients, it is more a container for any kind of data that was captured in relation to the patient, directly or indirectly, during an encounter between a patient (subject) and an investigator. One important difference is that CDISC-ODM multi-language support: one and the same item/question can be translated into different languages for use at different sites in different countries and cultures. In HL7-FHIR, this requires several instances of "Questionnaire", one for each language.
Patient	SubjectData	In ODM, "SubjectData" is a container for all data points about a specific patient (traditionally designated as "subject"). ODM however itself does not define what these data points are. In HL7 however, the "Patient" resource exactly describes the primary properties of the patient such as date of birth, sex, etc..
ClinicalImpression	SubjectData	HL7 ClinicalImpression describes "A record of a clinical assessment performed to determine what problem(s) may affect the patient and before planning the treatments …". Part of the ODM "SubjectData" can indeed be mapped to this resource, when describing the state of the patient before study start.
EpisodeOfCare	StudyEventData	The HL7 resource "EpisodeOfCare" can be mapped to a set of encounters between a patient and a care provider. As such it can also be used to describe all clinical study data that were captured as a result of a number of encounters (ODM "StudyEvents") between a subject and an investigator.
QuestionnaireResponse	FormData	The HL7 "QuestionnaireResponse" is limited to answers of "questions" filled when responding to a questionnaire. ODM however does not distinguish between data that comes from questionnaires only, but also data that come from other sources, such as lab data. Originally, ODM "FormData" was indeed essentially meant for paper forms, but as more and more electronic data capture (EDC) became available, it was promoted to a container for data that belong together and that were captured in one way or another, directly or indirectly during an encounter between subject and investigator.

This list does not comprise any of the CDISC-ODM extensions, such as the "Study Design Model in XML" (SDM-XML) [30], where the Element "ActivityDef" nicely maps to the FHIR resource "ActivityDefinition". Other SDM-XML elements such as "Workflow" can also be mapped to FHIR resources from the "Workflow" resource group. Leroux et al. also proposed new FHIR resources: ClinicalStudyPlan and ClinicalStudyData [29], which are the FHIR equivalents of the ODM "Study" (study definition) and "ClinicalData" (captured data) elements. These two new proposed FHIR resources should make the mapping with CDISC ODM complete, at least at the semantic level.

Mappings are nice, but data transformations should be avoided whenever possible as they inherently lead to information loss and can easily lead to errors [31,32]. Therefore, it is a good idea to also let the ODM standard evolve towards the HL7-FHIR standard.

The CDISC ODM development team recently started working on a new generation ODM standard. The name of the project is "ODMv2". Requirements were developed and can be summarized as:

- Backwards compatibility as much as possible. If this cannot be guaranteed, an XSLT stylesheet should be delivered that transforms ODM version 1.x into ODM version 2
- Support for alternative formats such as JSON. This probably means that the XML implementation will remain "leading", as many of the rules of the standard are implemented by means of XML-Schema and Schematron.
- A standardized RESTful web services API. A number of vendors [33,17] already have developed RESTful web services for exchange of ODM data and metadata. A standardized RESTful web services based API would also allow to work with distributed data, i.e. that all the data points of a single subject do not necessarily need to reside on a single server, but can be obtained through a set of RESTful web service queries. This especially becomes important when data from electronic health records is used.
- More flexibility in study design. The HL7-ODM standard originates from the times that most of the data was captured on paper forms, or EDC "CRF screens" at the best. This paradigm is outdated. Not all data is collected in "forms" and during "visits". There is a strong tendency to "e-Source" where data can come directly from the hospital information system, from electronic health records (that can use FHIR resources) and from devices such as wearables. Even more, "virtual" or "remote trials" become more common [34] where the subject (almost) never visits the clinic or has encounters with the investigator. This means that the hierarchy "visit - form - item group - item" must be revised.
- Better support for multiple controlled terminologies. CDISC-ODM only has the concept of "codelist" and "external codelist". The former defines value lists defined by the designer of the study (e.g. sponsor-defined lists for possible answers to a question) or copied from CDISC controlled terminology [26], the latter to coding systems such as SNOMED-CT, LOINC and others. However, even the names of these external codelists is not standardized. HL7-FHIR has a better mechanism for this [35]. Also in ODM, it is currently not possible to create subsets of external codelists, for example stating that a SNOMED-CT term or code should be used, but only from a selection of these.

These requirements are such that it will not be possible to guarantee 100% backwards compatibility. This can however be overcome by providing an XSLT stylesheet that

allow to transform ODM v.1.x in ODM v2 documents. Support for JSON also means that "roundtripping", i.e. transforming an ODM document to JSON and then transforming it to XML again without any information loss, must be possible. Whereas FHIR had the advantage of starting from scratch and has a very good mechanism for this, this will not be so easy for ODM.

Within the scope of ODMv2, it will not yet be possible to make a complete move to "resources". The reason is that the clinical research world is very conservative, and vendors of EDC systems will highly probably not want to invest in data capture tools that use a completely different approach. However, in many cases, a resource can be implemented as an ODM "ItemGroup", by grouping items that logically belong together and annotating them with codes and designations from the healthcare world. For example, an ODM "ItemGroup" "blood pressure" or even "vital signs" essentially corresponds to HL7-FHIR profiles "blood pressure" [36] and "vital signs" [22]. So, by semantically "standardizing" concepts and putting them in a single ODM "ItemGroup" and/or "Form" and annotating them with codes from the healthcare world (in the case of "blood pressure" using LOINC) and/or CDISC controlled terminology codes, makes them resemble FHIR resources and profiles already, and allows to exchange definitions of "biomedical concepts" [24].

A major obstacle for coming to a single data standard however remains the difference in the semantic standards used. HL7-FHIR tries to use existing semantic standards as much as possible that are well accepted in healthcare, but that were not necessarily developed within HL7, such as SNOMED-CT and LOINC. These semantic standards usually are pre-coordinated, i.e. they combine different pieces of information into a single term or code. For example, the LOINC code 1751-7 describes the test "Albumin [Mass/volume] is Serum or Plasma". CDISC however only uses controlled terminology to be used in post-coordination. For example, the code "ALB" describes "Albumin", which then needs to be combined with other terms to come to a description of a quantitative albumin test in blood. The background of this is that it remains (unfortunately) very unusual to exactly describe test to be performed in clinical study protocols. Even the CDISC Therapeutic Area Guides (TAUGs) [37], describing best practices for performing clinical studies in specific therapeutic areas, do not exactly describe recommended specific tests to be performed. Instead, very unspecific wording is used such as "test glucose in urine". It is then up to the sponsor to translate this into more specific tests, but even then, LOINC and SNOMED-CT coding is very seldom used. Upon submission to the regulatory authorities, the information is then post-coordinated (categorized) using CDISC controlled terminology. This means that for each test, the sponsor assigns a group of codes to a single test, for example for the compound, the specimen type, the test method, etc. The CDISC standard for doing so is the Submission Data Tabulation Model (SDTM) [38], providing a transformed "tabular view" on the data, and containing as well derived data as having data redundancy: essentially moving the data from operational data (ODM) to SDTM is an Extract-Transform-Load (ETL) process. As stated, SDTM requires CDISC controlled terminology (post-coordinated) to be used which is incompatible with healthcare coding systems such as LOINC and SNOMED. Also a "decomposition" from LOINC and SNOMED-CT codes (among others) has been shown to be very problematic [39]. In SDTM, use of these healthcare codes is very limited, or even not allowed (e.g. UCUM notation for units). A further complication is that the regulatory authorities still require data to be submitted in a completely different, but outdated format, named SAS Transport 5 format (often referred to a "XPT format). This is a binary format from the era of mainframe computers.

It has severe limitations restricting variable names to 8 characters, labels to 40 characters and values to 200 characters [40]. Furthermore, it only allows ASCII characters. The CDISC-XML team developed an XML standard for such tabular data, named Dataset-XML [41] but it is still not accepted by the regulatory authorities. This seriously delays the acceptance of XML as a technology within the clinical research world. Unfortunately, SDTM as a semantic standard is strongly influenced by the constraints of the XPT format, such as variable names to be constraint to 8 characters. HL7-FHIR does not have any such restrictions. Dataset-XML as a format does not have these constraints either, but for some of the information, instance files still need to implement these constraints for data when the dataset is for an SDTM submission. As long as these constraints exist, integration with healthcare data may remain difficult, if not impossible.

4. Discussion

This article describes some of the differences between the HL7-FHIR standard for exchange of data in the healthcare world with the CDISC-ODM standard for exchange of data in the clinical research world. Recent efforts to create new FHIR resources for use in clinical research are described, as well as the current efforts of the CDISC-ODM development team to modernize the ODM standard and add a number of features that are already supported by HL7-FHIR or even form the basis of HL7-FHIR. Unifying HL7 FHIR and CDISC ODM does not seem to be possible at this moment, but due to serious efforts of volunteers both from the HL7 side as from the CDISC side, such a unification may well be possible in not too far future.

References

[1] HL7, http://www.hl7.org. last access, January 2018.
[2] H. Oosterwijk, Will CDA replace HL7 version 2 messaging? http://www.auntminnie.com/index.aspx?sec=ser&sub=def&pag=dis&ItemID=104003. Last access: January 2018.
[3] W. Goosen, L.H. Langford, J.Am.Med.Inform.Ass. 21 (2014), 363-368.
[4] T. Benson, The HL7 V3 RIM. In: Principles of Health Interoperability HL7 and SNOMED. Health Informatics (2010). Springer, London.
[5] FHIR Release 3. https://www.hl7.org/fhir/, last access: January 2018.
[6] M. Khalilia, M. Choi, A. Henderson, S. Iyengar, M. Braunstein, and J. Sun, AMIA Annu, Symp, Proc. (2015); 717–726.
[7] FHIR Release 3 - Formats. https://www.hl7.org/fhir/formats.html. Last access: January 2018.
[8] Operational Data Model. https://www.cdisc.org/standards/transport/odm. Last access: January 2018.
[9] S.Hume, J.Aerts, S.Sarnikar and V.Huser, J. Biomed. Inform. 60 (2016), 352–362.
[10] FDA Technical Conformance Guide UCM384744, available at: https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf.
Last access: January 2018.
[11] Dataset-XML. https://www.cdisc.org/standards/transport/dataset-xml. Last access: January 2018.
[12] M.R. Cowie, J.I. Blomster, L.H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, M. Leenay, A. Michel, S. Ong, J.P. Pell, M.S. Southworth, W.G. Stouth, M. Thoenis, F. Zannad, A. Zalewski, Clin. Res. Cardiol. 106 (2017), 1–9.
[13] S. Bellary, B. Krishnankutty, and M.S. Latha, Perspect. Clin. Res. 5 (2014), 159–166.
[14] FHIR Publication Directory. https://www.hl7.org/fhir/directory.html. Last access: January 2018.
[15] FHIR Maturity Model. http://wiki.hl7.org/index.php?title=FHIR_Maturity_Model. Last access: January 2018.
[16] CDASH. https://www.cdisc.org/standards/foundational/cdash. Last access: January 2018.

[17] OpenClinica REST services. https://docs.openclinica.com/3.1/technical-documents/rest-api-specifications. Last access: January 2018.
[18] Medidata RESTful APIs. http://developer.imedidata.com/desktop/Overview/RESTful_APIs.htm. Last access: January 2018.
[19] FHIR RESTful API. https://www.hl7.org/fhir/http.html. Last access: January 2018.
[20] FHIR Resource Bundle. https://www.hl7.org/fhir/bundle.html. Last access: January 2018.
[21] Profiles defined as part of FHIR. https://www.hl7.org/fhir/profilelist.html. Last access: January 2018.
[22] FHIR Vital Signs Profile: https://www.hl7.org/fhir/observation-vitalsigns.html. Last access: January 2018.
[23] LOINC. https://loinc.org/. Last access: January 2018.
[24] Biomedical Concepts and SHARE. https://www.cdisc.org/standards/share/biomedical-concepts. Last access: January 2018
[25] D. Iberson-Hurst, Phuse Annual Conference (2015), Vienna.
[26] CDISC Terminology. https://www.cancer.gov/research/resources/terminology/cdisc. Last access: January 2018
[27] Unified Code for Units of Measure. http://unitsofmeasure.org/. Last access: January 2018.
[28] S.J. Steindel, J. Am. Med. Inform. Assoc. 17 (2010), 144-147
[29] H. Leroux, A. Metke-Jimenez, and M.J. Lawley, J. Biomed. Semantics (2017), 41-54.
[30] Study/Trial Design Model in XML. https://www.cdisc.org/standards/transport/sdm-xml. Last access: January 2018.
[31] A Manifesto for The Next Generation of Clinical Research Data Standards. W. Kubick, https://waynekubick.com/2016/08/30/cr-3-0-a-manifesto-for-the-next-generation-of-clinical-research-data-standards/. Last access: January 2018.
[32] V. Huser, C. Sastry, M. Breymaier, A. Idriss, J.J. Cimino, J. Biomed. Inform. 57 (2015); 88–99.
[33] Using web service technologies for incremental, real-time data transfers from EDC to SAS. A.Newbigging, Phuse 2010 (2010), Paper AD08.
[34] S. Jadhav, Virtual Clinical Trials: The Future of Patient Engagement? Appl. Clinical Trials, July 2016, http://www.appliedclinicaltrialsonline.com/virtual-clinical-trials-future-patient-engagement. Last access: January 2018.
[35] FHIR Code Systems. https://www.hl7.org/fhir/terminologies-systems.html. Last access: January 2018.
[36] FHIR Structure Definition Blood Pressure. https://www.hl7.org/fhir/bp.html. Last access: January 2018.
[37] CDISC Therapeutic Area Guides. https://www.cdisc.org/standards/therapeutic-areas. Last visited: January 2018.
[38] CDISC Study Tabulation Model. https://www.cdisc.org/standards/foundational/sdtm. Last visited: January 2018.
[39] J. Aerts, Eur. J. Biomed. Inf. 11 (2015), en2.
[40] C.I. Matthews, B.C. Schilling, Validating Clinical Data reporting with SAS, The SAS Institute, 2008.
[41] Dataset-XML. https://www.cdisc.org/standards/transport/dataset-xml. Last access: January 2018.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-64

Cross-Enterprise Communication and Data Exchange in Radiology in Austria: Technology and Use Cases

Emmanuel HELM[a,1], Andreas SCHULER[a] and Herwig MAYR[b]
[a]*Advanced Information Systems and Technology Research Group, University of Applied Sciences Upper Austria, Hagenberg, Austria*
[b]*School of Informatics, Communications and Media, University of Applied Sciences Upper Austria, Hagenberg, Austria*

Abstract. *Background*: The research project REPO (Radiology Ehealth PlatfOrm) was started in 2017 with the goal to "enable cross-enterprise collaboration in radiology using the Austrian eHealth infrastructure". *Objectives*: The objective of this paper was to provide an overview of the radiology IT environment – actors, use cases and technology. *Methods*: We conducted semi-structured expert interviews with radiologists and hospital operators and we statistically analyzed the client database of our research project partner. *Results*: Interviews led to a list of use cases where cross-enterprise collaboration in radiology takes place and the data analysis provided insights on the systems, networks and standards in place. *Conclusion*: The Austrian IT infrastructure in radiology is a heterogeneous naturally grown environment. Future developments should be based on internationally accorded standards and on integration profiles provided by Integrating the Healthcare Enterprise (IHE).

Keywords. Radiology, Health Information Interoperability, Electronic Health Records

1. Introduction

"Why is it so difficult to build IT systems that support a seamless flow of information along healthcare processes?" [1]

Insufficient communication and missing information in medical management are among the major factors contributing to *adverse events*, i.e. unintended injuries caused by medical management rather than the disease process [1].

To improve this situation and thus the quality of care, the FFG[2]-funded research project REPO (Radiology Ehealth PlatfOrm) was started in 2017 with the goal to "enable cross-enterprise collaboration in radiology using the Austrian eHealth infrastructure". The project aims at the cross-organizational processes in radiology, i.e. the communication of patient related information, reports and images, between different organizations. Project partner is CAS[3], national market leader for RIS (Radiology Information System) in practice settings.

[1] Corresponding Author: Emmanuel Helm, Softwarepark 13, 4232 Hagenberg, Austria, E-Mail: emmanuel.helm@fh-hagenberg.at
[2] The Austrian Research Promotion Agency
[3] Computer Anwendungssysteme GmbH, www.cas.at

In order to take the first steps to reach the goal, this paper aims to answer the following questions providing the basis for future research and development: (Q1) Who are the actors involved in these processes? (Q2) What are the use cases, i.e. what is the purpose of their communication? (Q3) What is the status quo, i.e. how do current solutions look like?

2. Methods

To answer the questions stated in the introduction we followed a three-step approach:

1. First, we determined a group of individuals that we expected to have in-depth knowledge of the processes.
2. Second, we conducted semi-structured expert interviews to identify the use cases and actors involved.
3. Last, we analyzed the state of the art of systems and technologies in place based on the interviews and by statistical analysis of our partner's client database. The interviews were transcribed and a narrative synthesis is provided in the results section below.

2.1. Finding Experts

The expertise of radiologists is per definition and based on their training in the area of the medical processes. However, in Austria the management of hospital radiology departments and the management of private radiology practices typically also lies in the hands of radiologists. This results in a group of individuals who understand both the organizational and the medical processes.

To cover both organizational worlds, hospital and practice, and thus all cross-organizational processes we decided to interview experts from both. In addition to the differences between hospital and practice, the Austrian federal healthcare system also leads to different approaches in different provinces [3]. Since our project partner services radiology practices in all nine provinces, we used those contacts to identify interviewees. Additionally, we contacted hospital operators to get in touch with their radiology departments.

2.2. Expert Interviews

A. Bogner and W. Menz define a typology of interaction situations and interview strategies [2]. We aimed at a situation where we act as "experts from a different knowledge culture" (type 2). This type indicates a symmetrical interaction situation where counterquestions from the interviewee are common. A precondition on the interviewer's side is the mastery of the specialist vocabulary. The semi-structured nature together with the situation of the interview lead to a dialogue-oriented, exploratory expert interview that is suitable for fact and data gathering [2].

2.2.1. Interview Guide

The interview guide is structured in three parts: (A) Data exchange, (B) Use cases and (C) Miscellaneous. Part (A) comprises four sectors: (A.1) *Hospital and Radiology Information Systems* (HIS/RIS), (A.2) *Picture Archiving and Communication Systems*

(PACS), (A.3) *Radiology Image and Report Communication Systems* and (A.4) *Use of Image and Report Communication Systems*. Part (B) is concerned with two use cases (B.1) *Expert Opinion* and (B.2) *Feedback*. Part (C) covers all topics and issues raised during the interviews that could not be related to A and B and is listed for documentation purposes.

(A.1), (A.2) and (A.3) aim at a listing of used systems and technologies in the interviewees' organizations. Moreover, those sectors allowed the interviewer to assess the interviewees' IT expertise. The questions usually started with "Which HIS/RIS are you using in your organization?" or "(How) do you use the Austrian electronic health record (ELGA)?". Depending on the course of the interview it deepened in details like "What is the bandwidth of your synchronous/asynchronous connections for image and report communication?" or "Who are the senders/receivers of your directed communication channels?". Another important aspect was to identify the share of non-digital communication, i.e. printed reports or images.

(A.4) aims to identify how and in which cases the Image and Report Communication Systems from (A.3) are used. The questions in this section range from "How often do other organizations/departments actively request reports/images per day?" to "*How* do you use *which* systems, to meet these requests?".

(B.1) follows the European Society of Radiology (ESR) definition of *Expert Opinion* as one type of teleradiology [4]: "*...usually occurs when specific in-house knowledge about a radiological subspecialty is insufficient or unavailable. When a radiologist with a specific expertise is consulted this should be called an expert opinion.*" The aim of (B.1) is to discuss this issue and find out *if* and *how* the interviewees manage the required report/image communication.

(B.2), *Feedback*, is about feedback loops from the recipients of radiology reports/images to the responsible radiologist, enabling quality management. We discussed with the experts what kind of outcome-quality management mechanisms are in place and how they could be improved.

For both (B.1) and (B.2) typical questions are "When does it make sense?", "What is the state of the art?" and "How could it work (better)?".

2.2.2. The Interviews

We conducted interviews with experts from four different hospital owners and with four radiology practice managers. We covered five Austrian states and several different types of organizations: university clinics, province hospitals, hospitals operated by a religious order, CT/MRT institutes and radiologists with radiography focus. To fit the interviews in the tight schedules of our interviewees, the duration aimed at 30 minutes.

2.3. Systems and Technologies

We expected that there are several different systems and technologies in place for report and image communication. To get an overview of the environment and answer question Q3 we analyzed the referenced systems from the interview (A.1-3) and the client database of our research project partner CAS. The database comprises basic information about the systems used by over 100 clients (exact numbers omitted due to business confidentiality). Moreover, we analyzed the current architecture documents regarding report and image communication via the ELGA [5].

3. Results and Discussion

This section represents a narrative synthesis of the expert interviews and the results of the statistical analysis of the CAS client database. It comprises the main insights structured in subchapters and discusses implications.

3.1. Who are the actors involved in these processes? (Q1)

Simply put, the actors of cross-enterprise communication of radiological reports and images are hospitals and practices. All combinations, hospital-hospital, hospital-practice and practice-practice are relevant.

The involvement of patients in the transport of reports and images was not uniform for the different interviewed organizations. A major finding of the interviews was that in Austria the ownership of the images is not clearly regulated (c.f. [6]). Thus, the handling of images differs greatly, from not giving the patient images to always handing the patient printed or recorded images.

3.2. What are the use cases, i.e. what is the purpose of their communication? (Q2)

During the interviews, different use cases for cross-organizational communication of radiology reports and images were identified. Based on the relative frequency of these use cases we separated them in primary and secondary use cases.

It is important to note that not every organization handles all of these use cases and that the relative frequency of the use cases differs between the organizations. Note: This list is most likely incomplete since it is based on a limited number of expert interviews.

3.2.1. Primary use cases

- UC1 – *Assignment*: Organization A assigns the patient to organization B, requesting a radiological examination to answer a medical question.
- UC2 – *Request*: Organization A requests existing reports and/or images from organization B.
- UC3 – *Referral*: Organization A refers a patient to Organization B. The patient and in some cases previous reports and/or images are transferred to B. In this case, A does not expect to receive further information from B.

Note that in Austria there is usually not a direct assignment (UC1) or referral (UC3) to a specific organization since the patient has the free choice of health professionals (*freie Arztwahl*), which has a positive effect on accessibility and patient satisfaction [7]. Thus A and B were chosen to emphasize the fact that they are distinct organizations, although the communication is not necessarily directed.

3.2.2. Secondary use cases

- UC4 – *Mamma-Screening*: "*Früh erkennen"* is an Austrian nationwide screening program for early detection of breast cancer [8]. It led to a new use case for cross-enterprise communication since the program obliges radiologists to always obtain a second opinion on the images. Organization A sends images for a specific case to organization B, requesting a BIRADS (Breast Imaging Reporting And Data System) assessment.

- UC5 – *Preliminary Reads* [4]: Typically occurring in emergency situations or due to seasonal peaks when there is no local radiologist available. Organization A requests an external radiologist from organization B to perform radiological image interpretation. The on-site radiologist usually composes the final authenticated report the next day during daytime office hours [4].
- UC6 – *Expert Opinion* [4]: A local radiologist or clinician in Organization A requests an expert opinion from a radiologist with specific expertise in Organization B.

3.3. What is the status quo, i.e. how do current solutions look like? (Q3)

3.3.1. Reports and Images

A major finding of the interviews was that associated reports and images are transmitted separately. For example, it is common for UC1 scenarios that the organization fulfilling the request of organization A sends the report via fax and the images via digital means. Organization A then has to re-establish the association between report and images.

There are also scenarios of use cases where the receiving organization is only interested in reports *or* images. While a general practitioner does not necessarily need a comprehensive CT study but the radiology report, an orthopedist might be more interested in the images than the report, performing the interpretation of the studies him- or herself.

3.3.2. Digital and Non-digital Communication

Each organization has a number of other organizations it has to communicate with in context of the use cases above. We found that bigger organizations (e.g. hospitals) dictate their means of communication to smaller organizations. For example some hospitals only accept non-digital reports (via fax or paper-based) while others insist on digital report communication.

All interviewees preferred to receive reports and images digitally to enable easier integration into their own information systems. In hospitals and practices alike, the scanning and assignment of printed reports from other organizations yields considerable effort and thus costs. The same goes for printed and recorded (CD/DVD) images.

3.3.3. Involved Systems and Technologies

There are lots of different data formats, networks and transmission technologies for requests, referrals, reports and images. A comprehensive listing of all these assets for all organizations in Austria is to our best knowledge not available and was out of scope of our interviews. However, to get an overview of the heterogeneous landscape, we added a description of our findings in the client database of CAS. Of course, these organizations use the RIS *CASmed*, which covers about 60 percent of all radiologists and CT/MR institutes in Austria (according to CAS). The other major vendors for RIS in Austrian practice settings are *Lukassoftware*[1] and *D.A.T.A. Corporation Sofwareentwicklungs GmbH*[2].

[1] http://www.lukassoftware.com/

[2] https://data.at/

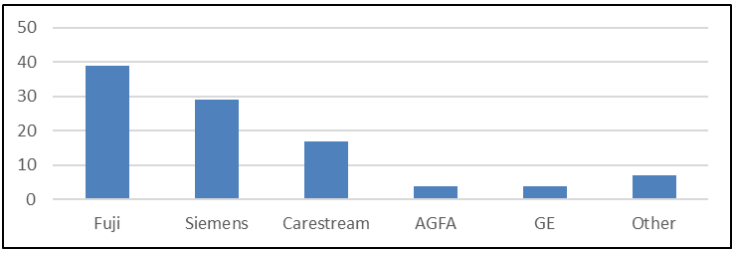


Figure 1. Distribution of PACS vendors in CAS clients.

From over 100 CAS clients, 85 percent use PACS from Fuji, Siemens or Carestream (see Figure 1). Overall, there are PACS from 12 different vendors in use and in some cases the software versions vary for one vendor.

For the digital communication of reports, there are two major vendors of solutions, A1 DaMe and HCS medical net. These solutions are also able to communicate, so you can send reports between them. Both are mainly based on the network of the e-card system GIN [9] but can also integrate clients from the internet.

While some PACS also support direct cross-organizational access to images (*Web PACS*), in general there are separate software modules for image communication. Several different solutions are used in CAS clients, e.g. Teleimage and HCS med vision. The networks for transport of DICOM studies vary strongly depending on the province and are often connected to or combined with a radiological long-term archive. Proprietary *Web PACS* solutions are most common in rural environments where the number of cooperating organizations is not too high.

The data formats used for digital reports are Edifact text, Edifact PDF and HL7 v3 CDA, with the majority (>90%) still based on the former two. Images are completely DICOM-based. Assignments and referrals are primarily paper-based and transferred postal, by fax or carried by the patient. In some settings, digital communication of assignments and referrals between certain organizations was established based on HL7 v2 messages.

3.3.4. ELGA

Radiology report communication based on the ELGA infrastructure is in the early stages and image communication is not supported yet [5]. As of February 2018, six clients of CAS register their radiology reports in ELGA. Hospitals are already obliged to register their reports since 2016.

During the interviews, we received positive feedback from radiologists in practice settings when they were able to view reports from cooperating hospitals. However, the need for more documents, e.g. reports from other specialties, was pointed out. An important issue that all interviewees in practice settings agreed on was the need for further integration of the workflows, i.e. electronic assignment and referral. In hospital settings, some interviewees wished for better integration of external ELGA documents in their existing workflows.

3.4. Expert Opinion

The consultation of external specialists was quite uncommon among our interviewees and was said to happen only a few times a year in average, also because of the lack of

ways to identify the experts and establish communication. Since most organizations, practices and hospitals employ a number of different specialists, the first place to seek an expert opinion is in-house. Hospital radiologists also pointed out that there is no standardized way to charge for giving an expert opinion.

All interviewees agreed on two fields where expert opinions can be very useful: (1) *Neuroradiology*, a very specialized field of radiology; (2) *Bone neoplasms*, due to their low incidence and high variance.

The interviewees also agreed on the need for seamless integration of expert opinion systems into their existing tools. While some radiologists prefer voice communication to discuss the case in detail, others assume textual communication sufficient.

Apart from hardly used *official* ways of requesting expert opinion, the interviewees referred to open platforms like Radiopaedia[1] that are used in some cases to get expert opinions. Moreover, and more frequently, *informal* ways of getting a quick feedback from peers include the use of online messenger services.

3.5. Feedback

Giving feedback to preceding external medical professionals in the process of treating a patient is not common. Currently it works via personal contacts and mostly by phone, e.g. a hospital surgeon tells an acquainted external radiologist that the initially diagnosed meniscus tear proved uninjured during surgery.

Most interviewees said that there should not be an automated mandatory feedback for contradicting diagnosis. One interviewee stated that gathering feedback should always be an obligation of the preceding, not the succeeding professional. However, all interviewees agreed that there should be a mechanism to actively request feedback, which means that e.g. a radiologist can place a *feedback requested* flag on a report so that succeeding medical professionals know about the preceding one's interest and can confirm or disprove a diagnosis. Again, the feedback mechanism should be fully integrated in the respective information systems and not require additional organizational effort.

4. Conclusion and Outlook

The introducing question was "*Why is it so difficult to build IT systems that support a seamless flow of information along healthcare processes?*". R. Lenz and M. Reichert provided elaborated answers to the question starting from a fundamental differentiation of organizational and medical treatment processes [1]. In our work, we focused the organizational processes, i.e. the transport and provision of patient-related information, reports and images, across organizational boundaries. Our interviews and analysis confirm the findings of R. Lenz and M. Reichert that despite well-accepted standards for data integration like HL7 and DICOM, healthcare applications are still far from plug and play compatibility [1].

The Austrian healthcare IT environment is heterogeneous and naturally grown, but the recent trend towards healthcare networks and integrated care, e.g. ELGA, led to an urgent need for integration of these systems. The recently increasing activities in imaging and workflow topics in the wake of the ELGA project has the potential to significantly

1 https://radiopaedia.org/

improve the communication between healthcare organizations, ultimately leading to lesser organizational efforts and an improvement of the quality of care. This work aims to provide a basis for future developments, giving an overview of the state of the art, use cases and the heterogeneous environment.

We conclude that future solutions to enable the cross-organizational communication of reports and images should consider the surrounding workflows, including referrals and assignments, and should be based on internationally accorded standards and on integration profiles provided by *Integrating the Healthcare Enterprise* (IHE).

Acknowledgements

The authors thank all experts participating in the interviews for their cooperation, providing insights to the Austrian (e)Health environment. Moreover, the authors thank the "*OÖ Gesundheitsfonds, Koordinationsstelle eHealth"* for valuable discussions and their help in identifying experts.

References

[1] Lenz, R., Reichert, M. (2007). IT support for healthcare processes–premises, challenges, perspectives. Data & Knowledge Engineering, 61(1), 39-58.
[2] Bogner, A., Menz, W. (2009). The theory-generating expert interview: epistemological interest, forms of knowledge, interaction. In Interviewing Experts (pp. 43-80). Palgrave Macmillan UK.
[3] Hofmarcher, M. M. (2013). Das Österreichische Gesundheitssystem. Vienna, 20, 2013.
[4] European Society of Radiology (ESR). (2014). ESR white paper on teleradiology: an update from the teleradiology subgroup. Insights into Imaging, 5(1), 1-8.
[5] Repas, S., Klostermann, A., Seerainer, C., Kuttin, O., Hell, J., Rauchegger, G. (2017). ELGA-Gesamtarchitektur v2.30. ELGA GmbH, https://www.elga.gv.at/, last Access: 30.01.2018
[6] Wallner, F. (2010). Das Eigentum am Röntgenbild. Recht der Medizin, 2010/93(4), 107. MANZ.
[7] Habimana, K., Bachner, F., Bobek, J., Ladurner, J., Ostermann, H. (2015). Das österreichische Gesundheitswesen im internationalen Vergleich. Studie im Auftrag des BMG. Wien, Gesundheit Österreich GmbH (GÖG).
[8] Jeitler, K., Semlitsch, T., Posch, N., Siebenhofer, A., Horvath, K. (2015). Breast cancer screening in Austria: key figures, age limits, screening intervals and evidence. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 109(4), 363-370.
[9] Schuegerl, Rainer (2014). The Austrian E-Card System - Experiences of a Nationwide eHealth Infrastructure. University of Amsterdam, Amsterdam Privacy Conference 2012 (APC 2012).

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-72

Facilitating the Information Exchange Using a Modular Electronic Discharge Summary

Kerstin DENECKE[a], Pascal A. DITTLI[a], Niveadha KANAGARASA[a], and Stephan NÜSSLI[a,1]
[a]*Bern University of Applied Sciences*

Abstract. *Background:* Discharge summaries are a standard communication tool delivering important clinical information from inpatient to ambulatory care. To ensure a high quality, correctness and completeness, the generation process is time consuming. It requires also contributions of multiple persons. This is problematic since the primary care provider needs the information from the discharge summary for continuing the intended treatment. To address this challenge, we developed a concept for exchanging a modular electronic discharge summary. *Methods:* Through a literature review and interviews with multiple stakeholders, we analysed existing processes and derived requirements for an improved communication of the discharge summary. *Results:* In this paper, we suggest a concept of a modular electronic discharge summary that is exchanged through the electronic patient dossier in CDA CH level 2 documents. Until 2020, all Swiss hospitals are obliged to connect to the electronic patient dossier. Our concept allows to access already completed modules of the discharge summary from the primary care side, before the entire report is entirely finalised. The data is automatically merged with the local patient record on the physician side and prepared for data integration into the practice information system. *Conclusion:* Our concept offers the opportunity not only to improve the information exchange between hospital and primary care, but it also provides a potential use case and demonstrates a benefit of the electronic patient dossier for primary care providers who are so far not obliged to connect to the patient dossier in Switzerland.

Keywords. Discharge summary, Medical documentation, Information exchange, Transition of care, Electronic patient record, Electronic health record

1. Introduction

The discharge summary is an important document used for information exchange in the transition from the hospital to primary care. It contains information on the patient's medical history, type, extent and results of the diagnostics and therapy measurements, the patient's prognosis and concrete recommendations for further treatment. It summarizes the course of disease and therapy. More specifically, the discharge summary is an essential source of information for the follow-up treatment. The recipients of the document such as rehabilitation clinics, general practitioners, physiotherapists or nursing houses have different information needs. Poor quality discharge summaries have been repeatedly demonstrated to lead to increased adverse events in patient care after discharge and to a need for re-hospitalization. Research has shown that there is a gap in

[1] Corresponding Author, Stephan Nüssli, Bern University of Applied Sciences, Höheweg 80, 2502 Biel, Switzerland, E-mail: Stephan.nuessli@bfh.ch.

the information chain after discharge from hospital [1]. Despite the ideal of a seamless handover, discharge summaries may not reach the general practitioner (GP) or may be compromised by significant delays, poor quality or illegibility [2]. However, a complete, accurate, and timely discharge summary can communicate important information to the GP, prevent adverse events, and reduce hospital readmission [3].

In this work, we are focusing on the GP and their needs with respect to the discharge summary for continuing the health monitoring and treatment of a patient discharged from a hospital. Since the discharge summary is also a legal document, the generation and signature process of the complete report is comprehensive and time consuming. The final report needs to be signed by the responsible hospital physician. At the same time, it is obvious that the GP needs the information immediately after discharge to continue the treatment appropriately. For this reason, a short summary with the most relevant information is currently sent right after discharge of the patient. However, this leads to double efforts since the GP has to check twice whether the data in his local patient record corresponds to the information from the discharge summary short report and final report. Another delay is caused by the paper-based format in which the report is currently transferred to the primary care provider.

A recent study in Germany demonstrated that the communication between hospital and primary care is mainly realised by standard mail. However, more than 35% of the primary care physicians and of the hospital physicians desire an exchange via secure E-mail. 40% of the hospital physicians would like to use an integrated information system [4]. To address these issues, we introduce the concept of a modular electronic discharge summary. Instead of sending unstructured PDF documents by E-Mail or printed by standard mail, reports will be available electronically and structured through the electronic patient record. With the adoption of the federal law on the electronic patient record in Switzerland, hospitals are obliged to join a certified community of patients until 2020. From then on, patients will have the opportunity to open an electronic patient record (EPD, electronic patient dossier) and access collected documents. The EPD is designed to create the technical prerequisites for healthcare professionals to have the relevant patient data at the right place at the right time. However Switzerland decided for an opt-in concept. That means that the participation is mandatory only for hospitals and nursing homes. For Patients and GPs the participation in the EPD is by choice. That is called “Doppelte Freiwilligkeit”. Therefore there must be a clear benefit for Patients and GPs to engage for the EPD.

As a result, GPs will be able to spend less time updating the patient's medical history and are able to focus more on their patients instead. In this paper, we identify the technical and semantic conditions that must be given to introduce a modular electronic discharge summary in Switzerland. In addition, a concept for the technical implementation is described. The basis of this work is the electronic patient record (ELGA) implemented in Austria, in which the reporting is already structured modularly.

2. Material and Methods

In this section, we summarize the methodology for concept generation and provide details on the eHealth initiative in Switzerland that concentrates on an electronic discharge summary.

2.1. Requirements collection

For concept generation, we collected requirements from a local GP and from the head of the medical informatics department of a hospital. More specifically, we collected information on the current procedure of information exchange between the hospital and the local physician in these two interviews. We also discussed our concept with them and considered the feedback for adapting the concept presented in this work.

Further, we talked to several companies that are distributing hospital information systems and discussed our concept with them. Another interview was held with a senior consultant of Triamed, a software producer that is developing information systems for primary care. That interview focused on possibilities to integrate a modular discharge summary into a practice information system (PIS).

2.2. IPAG and eAustrittsbericht

The "eAustrittsbericht" suggested in December 2015 by an inter-professional working group in Switzerland (IPAG) formulates recommendations for the interdisciplinary use of the information modules that are most important during treatment transitions [5]. These are to be used in electronic documents of the "Transition of Care" (eToC), i.e. transfers and treatment transitions, for inter-professional communication. These data modules can be used independently of each other in various documents such as discharge summaries. An eToC document consists of four main modules: "Problems", "Treatments", "Medication" and "Recommendations and other measures". eToC documents are intended for all processes in which one or more treatment transition takes place.

The contents of the respective modules can be structured or coded, but entering free text should always be possible. In the discharge summary, we can find different sections that contain data originating from the various professional groups. For example, the nurses summarize the nursing interventions and the health status of a patient as well as educations the patient received. Dietitians describe nutrition assessments and diagnosis as well as nutrition related treatments. For each module, the professional group must be included in the data entry so that it is clear from which area the information originates (e.g. laboratory, nursing, physician, radiology). The objective is to enable the use of subject-specific or occupational group-specific designations. Whenever possible, the SNOMED CT reference terminology should be used as a nomenclature for coding, and only if this is not possible, data should be coded with another classification system or ontology.

The type of transfer report or eToC document determines which modules should be included. This is implemented with predefined options. IPAG opted for the following options: "Mandatory", "Recommended", "Possible" and "Not applicable". If an entry is mandatory, a specification must be made (zero values are not possible) and if no information is available, it must be specified why the information does not exist.

2.3. HL7 CDA

Like DICOM, HL7 CDA is a standard used for healthcare communications. A CDA document is saved in XML (Extensible Markup Language) format. CDA documents can be structured according to three different levels. For CDA Level 1, only the header must be structured. The body contains unstructured textual information. It is also possible to

attach PDF files, TIFF images or other documents. In contrast, the body of a CDA Level 2 document must contain structured data so that the content can be recognized. In Level 2, however, it is also possible to enter unstructured data in the body. The Swiss healthcare system adopted the CDA standard by adding Swiss specific requirements. This resulted in the standard CDA-CH, which is used for the Swiss EPD. Meanwhile, the CDA-CH standard is already available in version 2 [6].

3. Results

In this section, we are describing our concept. We are focusing on the information exchange and information processing in hospitals and GPs. Hospitals have been chosen because they are legally obliged by the EPD law in Switzerland to join a certified community and participate in the EPD by 2020. GPs have been selected as a second stakeholder group because they are central to hospitals and the implementation of the EPD. The overall concept foresees a modular electronic discharge summary that is made available in the EPD and can be accessed and downloaded by the GP.

3.1. Requirements

Along with the discharge of a patient, two reports are generated. The short report and the final discharge summary differ mainly in scope and purpose. The short report is sent directly to the GP when the patient leaves the hospital and provides a rough overview of the most important key data such as diagnoses, allergies, procedures and medication. The final discharge summary additionally includes results of findings as well as follow-up documentation written by the resident physicians. The short report is usually only signed by the resident physician, but the final discharge summary contains a signature from the chief physician. Consequently, the concept of the electronic discharge summary therefore also requires a modular short report, unless the results of the downstream processes are sent separately.

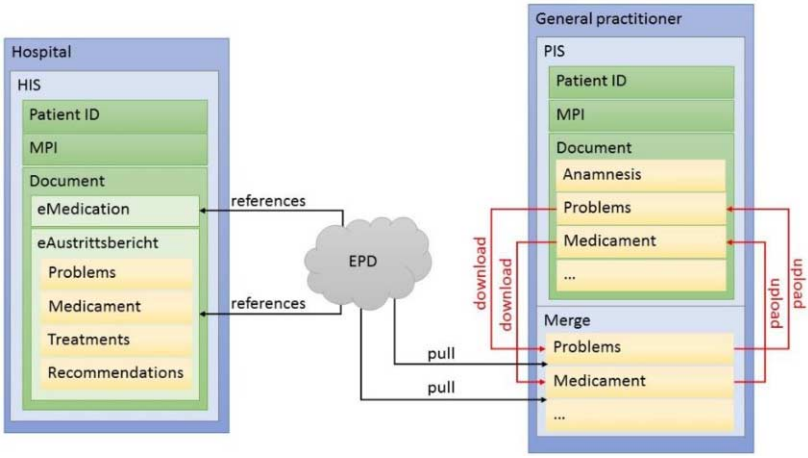


Figure 1. System architecture.

The data of the discharge summary should be made available through the EPD, thus, the information system has to be able to connect to the EPD. Data from the report should be stored locally in the patient record in the PIS. Interactions with the EPD and data updates in the PIS have to be logged in a logbook. A merging function is requested to see updates of the data such as medication data as reported in the discharge summary and to enable a comparison of the local data with the data from the report. Duplicates should be removed automatically. Redundant medications have to be made visible. The comparison has to be semantically, i.e. medications should be compared based on their active ingredients.

3.2. Modular structured discharge summary and its transfer

We introduce a modular electronic discharge summary. For each module of the modular electronic discharge summary, a status and the signature of the author for each module has to be assigned. In this way, the GP can see whether a module is already finalised or changes can still be made. Since the short report is usually targeted by the assistant doctor, the employment is included along with the signature. The head or chief physician is enabled to approve the final report at once, with the approval adopted for all modules. Signing individual modules requires that the different health professionals possess the necessary authorizations and work with their health professional ID. Once all modules are signed, the complete discharge summary can be imported by the GP.

On the primary care side, the access to the modular discharge summary is realized through a merging plugin in the PIS. In the background, the connection to the EPD is established and the data is loaded into the merge view. At the same time, the data is retrieved from the GP's patient record stored in the PIS. In addition, a copy of the original data from the EPD is stored in the local PIS. All executed actions are recorded and listed in a logbook.

The GP is informed as soon as all required modules contained in the discharge summary have been signed and approved. Afterwards, he can load the modules into the PIS and open the merge view. The data is retrieved from the referenced storage location as CDA-CH elements from the EPD according to the IHE profile XDS.b. In the merge view, the information contained in the patient record is compared to the information contained in the discharge summary pointing him to differences. After checking and confirming, the data in the PIS is updated. Figure 1 below illustrates the system architecture with the systems involved. The assumption is made that the GP's PIS is already connected to the EPD and thus modules from the hospital can be viewed directly.

PIS/Pat/Medication		Merge		KIS/Medication of Discharge	
Supplement	**Ingredient**	**Supplement**	**Ingredient**	**Supplement**	**Ingredient**
Beloc Zok	Metoprolol	Beloc Zok	Metoprolol	Dafalgan	Paracetamol
Norvasc	Amplodipin	~~Dafalgan~~	~~Paracetamol~~	Voltaren	Diclofenac
Sandoz	Paracetamol	Norvasc	Amplodipin	Beloc Zok	Metoprolol
Berocca	Vitamin B	Voltaren	Diclofenac	Norvasc	Amplodipin
		Sandoz	Paracetamol		
		Berocca	Vitamin B		

Figure 2. Merge view on the primary care provider side.

Figure 3. Signing of modules of the discharge summary.

Consider the following example: The medication is shown in the merge view (see figure 2). Duplicates are recognized and joined together during the joining process. For example, the drug Beloc Zok (Metoprolol) is only transferred to the merge view once, since it is present in the patient record and in the medication at discharge. If the two lists contain drugs with different names, but the same active ingredients, they are marked in red. Hospitals often use different medications than GPs, so it is quite possible that different medications with the same active ingredients may appear in the list. In this case, the GP can choose which medication he or she wants to have registered in the patient record. In green, the merge view shows the medications that were additionally distributed when the patient left the hospital. They can be integrated directly by the GP into his patient record after the checking.

4. Discussion

In this paper, we introduced a concept of a modular discharge summary that can be integrated into the eHealth environment of Switzerland. Switzerland's EPD is based on conceptual ideas similar to those of Austria. The Austrian health system has opted for an opt-out strategy, which means that an electronic patient record is created for all citizens by default. The ELGA portal offers the exchange formats eMedication and eBefund, which are used in the discharge summary. For a structured document, the Austrian HL7 group specified CDA, whereby the documents are subject to an XML structure. The CDA Release 2.0 is used, whereas Switzerland uses the same CDA with national components, namely CDA-CH V2. In the ELGA, eMedication and eBefund can be uploaded and managed on the portal. Through the portal, the patient can view his or her documents. From a technical point of view, CDA-CH allows the implementation of information modules, but it is recommended to use a CDA Level 2 structuring in which the body is also structured. The implementation of a national standard for the transmission of medical data would represent a milestone in terms of interoperability for the Swiss healthcare system. In this work, we did not design the concrete content of the CDA discharge document, but relied upon the general modular structure as suggested by the

inter-professional working group. Robust methods for reaching consensus among the relevant professional groups across all settings and all regions are paramount for the success of the introduced concept.

The idea of a modular discharge summary is not new. Paterson introduced their CDA-based structured discharge summary system in 2002 [7]. In Germany, the HL7 working group defined a structure for an eArztbrief (electronic discharge letter) [8]. It considers the developments in Austria and Switzerland. Physicians and physiotherapists can send an eArztbrief directly from the PIS. The data is transferred by a special secure mail provider. When incomplete letters are sent, this can be indicated by a status sign such as “preliminary”. Schabetsberger et al. developed and realized a strategy for a stepwise replacement of the paper-based transmission of medical documents with a distributed, shared medical record [9]. An electronic communication of discharge letters between existing information systems of different health care providers in Tyrol, Austria, has been established in the form of cryptographically signed S/MIME e-mail messages and, via a secure web portal system. Our work differs from this work by introducing a merge view that facilitates the integration of data for the GP and by directly integrating the EPD for data exchange instead of relying upon e-mail. Beyond, we did not focus on the HL7 CDA structure itself, but on how to use the components and integrated them into an eHealth environment.

An introduction to the use of modular discharge summary could be interesting for GPs because they would always be up to date in terms of information technology and could transfer the information directly from the modules to the patient's medical history. As many hospitals are competing with each other, the use of a modular discharge summary can make it easier for GPsto work together with a hospital and thus, should become an integral part of a hospitals' business strategy. In contrast to Austria, Switzerland has spoken in favour of an opt-in method, which means that patients have to make an effort on their own initiative to open a patient dossier. Due to this fact, Switzerland's electronic patient dossier must be convincing. Not only patients need to be convinced of the benefits of the EPDs, but also the GPswho have not yet been obligated to use the EPD by law. The future of the EPD therefore goes hand in hand with an added value for all, patients, hospitals and primary care provider.

Modular data exchange has the potential of shaping an interdisciplinary data flow. With the introduction of the EPD, reporting will also be revolutionized in a later phase. To be successful, the implementation of a modular exchange of discharge information needs to be coordinated by a central body such as eHealth-Suisse and the Federal Office of Public Health, since it involves all health professionals. The basic building block for implementing modular reports is the structuring of documents. Before a restructuring of the reporting system can be envisaged at national level, the large health care institutions must first be connected by the EPD and use the same exchange format.

As soon as reports can contain structured data and PIS can be connected to the EPD, the implementation of a merge view for general practitioners at the PIS manufacturers will be added to the roadmap. We hope that solutions such as the merge view will enable more physicians to be convinced of the benefits of electronic patient files in the future.

References

[1] C. Walraven, M. Taljaard, CM. Bell, E. Etchells, IG. Stiell, K. Zarnke, AJ. Forster. A prospective cohort study found that provider and information continuity was low after patient discharge from hospital. J Clin Epidemiol 2010;63:1000–10.

[2] S. Kripalani, F. LeFevre, CO. Phillips, et al. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007;297:831–41

[3] C. Walraven, R. Seth. Effect of discharge summary availability during post-discharge visits on hospital readmission. J Gen Intern Med 2002;17:186–19.

[4] K. Obermann, P. Müller, S. Woerns (eds.): Ärzte im Zukunftsmarkt Gesundheit 2017 Transsektorale Zusammenarbeit. Eine deutschlandweite Befragung niedergelassener Ärztinnen und Ärzte. January 2018, https://www.stiftung-gesundheit.de/pdf/studien/Aerzte_im_Zukunftsmarkt_Gesundheit_2017.pdf

[5] D. Aronsky, E. Bättig, S. Bürki Sabbioni et al. 2. Bericht. Interprofessionelle Arbeitsgruppe Elektronisches Patientendossier. Arbeitsgruppe eAustrittsbericht. Bern, 2015, https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/2016/D/160119_eAustrittsbericht_Bericht_IPAG_D.pdf

[6] J. Bleuer et al. CDA-CH-II: Spezifikation zum Erstellen von Vorlagen für die Health Level 7 Clinical Document Architecture. Version 1.2, HL7 Benutzergruppe Schweiz, 2011, https://www.hl7.ch/default/assets/File/Publikationen/CDA-CH-II_de_V1_2.pdf

[7] G.I. Paterson, M. Shepherd, Xiaoli Wang, C. Watters, D. Zitner: Using the XML-based Clinical Document Architecture for exchange of structured discharge summaries. System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on. 2002

[8] KU. Heitmann, F. Oemig, D. Hellmuth, M. Aschhoff: Implementierungsleitfaden Arztbrief 2014/2015 (Implementation guide Letter to the doctor), final version, 15.10.2015, http://download.hl7.de/documents/cdar2-arztbrief/Arztbrief2014-v100.pdf

[9] T. Schabetsberger, E. Ammenwerth, S. Andreatta, et al. From a paper-based transmission of discharge summaries to electronic communication in health care regions. Int J Med Inform. 2006 Mar-Apr;75(3-4):209-15. Epub 2005 Aug 22.

[10] H.B. Bludau, A. Wolff, A.J. Hochlehnert, Presenting XML-based medical discharge letters according to CDA, Methods Inf. Med. 42 (5): 2003: 552—556

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-80

Effectiveness of Anonymization Methods in Preserving Patients' Privacy: A Systematic Literature Review

Mostafa LANGARIZADEH[a], Azam OROOJI [a,1], Abbas SHEIKHTAHERI[a]
[a] *Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran*

Abstract. Background: An ever growing for application of electronic health records (EHRs) has improved healthcare providers' communications, access to data for secondary use and promoted the quality of services. Patient's privacy has been changed to a great issue today since there are large loads of critical information in EHRs. Therefore, many privacy preservation techniques have been proposed and anonymization is a common one. Objectives: This study aimed to investigate the effectiveness of anonymization in preserving patients' privacy. Methods: The articles published in the 2005-2016 were included. Pubmed, Cochrane, IEEE and ScienceDirect were searched with a variety of related keywords. Finally, 18 articles were included. Results: In the present study, the relevant anonymization issues were investigated in four categories: secondary use of anonymized data, re-identification risk, anonymization effect on information extraction and inadequacy of current methods for different document types. Conclusion: The results revealed that though anonymization cannot reduce the risk of re-identification to zero, if implemented correctly, can manage to help preserve patient's privacy.

Keywords. Anonymization, confidentiality, electronic health record

1. Introduction

Today, healthcare systems have an increasing emphasis on modernizing infrastructures and replacing paper-based medical records with electronic health record systems (EHRs). This transformation makes new opportunities for secondary use of clinical data [1]. An EHR is a system embracing patients' demographic, diagnostic, lab and medication data. These data can be accessed and shared through computer networks among healthcare providers in different organizations [2]. Therefore, EHR is a rich resource for secondary uses such as research, quality assessment and epidemiology [2,3]. However, the high volume of identifiable personal data in EHRs would threaten individuals' privacy [4]. Studies showed that 59% of patients believed that EHR has increased risk of data loss or privacy breach [5]. In this regard, policies and regulations have been developed worldwide to restrict identifiable data sharing and to reduce privacy concerns, such as General Data Protection Regulation (GDPR) was approved by the European Union parliament on April 2016[6]. In the U.S., Health Insurance Portability and Accountability Act (HIPAA) enacted in 1996 and Health Information Technology for Economic and Clinical Health (HITECH) Act enacted in 2009 [7]. According to many regulations, the

[1] Corresponding Author: PhD candidate in medical informatics, Department of Health Information Management, School of Health Management and Information Sciences, No. 6, Rashid Yasemi Av., Vali-e-Asr St., Tehran, Iran. Tel: +98-21-88794301; E-mail: orooji.a@tak.iums.ac.ir

medical team, and those permitted by the patient and those authorized according to law are permitted to access patient's information [8].

Based on many of these policies, researchers and other secondary users can only access identifiable data only if they obtain the required permission from ethical committees and informed consent from patients. This procedure is time-consuming and sometime impossible, especially when the target population is large. However, if the data is anonymized, there is no need for informed consent [9-10]. ISO defines anonymization as the act of eliminating the links between identifiable data and the data subject [11]. In recent years, many techniques have been developed for privacy preservation for healthcare data such as Anonymization, Perturbation, Condensation, Randomization and Fuzzy based methods; among them, anonymization showed to be a promising method [12-13]. Although, anonymization techniques are capable of preserving privacy, they may negatively affect data utility for secondary uses. [12]. In fact, anonymization methods and the resulted anonymized data may result in negative effect on secondary use of data. Hence, many studies have been conducted to investigate these issues and develop solutions for the problems. The aim of this research was to systematically review and categorize the problems of the anonymization techniques and their effect on secondary use of patients' data.

2. Methods

The articles published in the 2005-2016 (conference and peer-reviewed papers) were included. Pubmed, Cochrane, IEEE and ScienceDirect were searched with a variety of related keywords (Table 1). Newspapers, reviews, letter to editor, workshop reports, posters, short reports, books and thesis, articles written in non-English language and also papers related to anonymization on non-health data were excluded. Articles with no access to the full-text were also excluded. In addition, articles were selected if they related to secondary use of anonymized data. In fact, papers related to develop methods for anonymization or those related to problems of these methods were also excluded.

A total of 659 papers were initially identified. All duplicated (n=58) and non-English (n=13) articles were removed using Endnote software, however, a manual revision was done for verification. Two reviewers independently screened titles (n=588) and abstracts (n=165) and then reviewed the full texts (n=46). Discrepancies resolved by consensus. Finally, 18 publications were included in the review. 28 papers after reviewing full text were excluded mainly because they were related to the problems of the methods not the problems of anonymized data for secondary use.

3. Results

Based on the included papers (aims and results), we classified and discussed the issues of anonymizations in four categories including: (1) data secondary use (SU), (2) re-identification risks (RR), (3) effect on information extraction (IE), and (4) Inadequacy of current methods for heterogeneous documents (IN). Following paragraphs describes each of these categories in detail.

Table 1: The search query used in this study

Science Direct	pub-date > 2005 and pub-date < 2016 and (TITLE-ABSTR-KEY(de-identif*) or TITLE-ABSTR-KEY(deidentif*) or TITLE-ABSTR-KEY(Anonymization) or TITLE-ABSTR-KEY(De-personalization) or TITLE-ABSTR-KEY(Depersonalization) or TITLE-ABSTR-KEY(Pseudonymization)) and ("electronic health record" or "electronic medical record").
Pubmed	(Electronic health record [All Fields] OR Electronic medical record [All Fields]) AND (de-identif*[Title/Abstract] OR deidentif* [Title/Abstract] OR Anonymization[Title/Abstract] OR De-personalization [Title/Abstract] OR Depersonalization [Title/Abstract] OR Pseudonymization [Title/Abstract]) AND ("2006/01/01"[PDAT] : "2016/01/01"[PDAT])
Cochrane	(electronic health record:ti,ab,kw or electronic medical record:ti,ab,kw)and (deidentif*:ti,ab, kw or deidentif*:ti,ab,kw or Pseudonymization:ti,ab,kw or anonymization:ti,ab,kw) Publication year from 2006 to 2016
IEEE	((electronic health record OR electronic medical record) AND (Abstract:deidentif* OR "Abstract":anonymization OR "Abstract":Pseudonymization OR "Abstract":deidentif* OR "Document Title":deidentif* OR "Document Title":anonymization OR "Document Title":Pseudonymization OR "Document Title":deidentif*)) and refined by Year: 2006-2016

3.1. Data secondary use

Secondary usage of health data plays a key role in promoting medical knowledge. In the primary use of EHRs, providing healthcare services, it is necessary to include patient's identification information within the records. In secondary use, however, there is no need for this information [14]. In recent years, there have been many techniques used to preserve patient's privacy. However, one of the most problem of these techniques is possible elimination of much valuable information required for the research purposes and other secondary uses [12]. Therefore, many scientific articles focused on this issue and even proposed a number of methods to strike a balance between privacy preservation and maintaining data value. In this regard, seven papers out of 18 included articles focused on this issue [9, 15-20] (Table 2). For example, Neuberger [18] investigated anonymization approaches and showed that pseudonymization was the best method of striking a balance between data secondary use and privacy preservation. As another example, applications need to be piloted before use in actual environment. Currently, testing is performed with fake data often leads to worse code coverage and fewer uncovered bugs, so testing with real data is important. However, different data privacy laws prevent organizations from sharing these data with test centers because databases contain sensitive information. In this regard, Grechanik [19] proposed a solution for use of anonymized real data in evaluating the effectiveness of such applications.

3.2. Re-identification risk

Re-identification is a process in which attempts taken to find the owner of a record or document which has already been anonymized [21]. Attackers can re-identify data by linking the anonymized data to the other accessible datasets; therefore, anonymization techniques do not guarantee the anonymity of data [16, 22]. Four studies out of included papers focused on this issue [20, 22-24] (Table 2). Some studies introduced methods of estimating the re-identification risk of records [22]. In spite of all efforts to prevent re-identification, the necessity of right legislation for the relevant delinquencies is explored by some researchers [23]. Only one investigation addressed the details of cost-effectiveness evaluation of re-identifying of health data. In this study, the cost of each record was estimated in accordance with the value of each attribute [20]. In addition, El-

Emam [3] found that many attacks succeed due to the inefficiency of the existing anonymization methods.

3.3. Effect on information extraction

Anonymization is a barrier to implementing effective data retrieval mechanisms. Since, it is not possible to do effective query for relevant data using anonymized data [25]. Due to the significance of the issue, many investigations have been conducted to assess the effect of anonymization on different operations such as data extraction or retrieval (nine papers) [2, 9-10, 12-13, 16-17, 25-26]. Some researchers proposed strategies to solve this problem. For instance, an efficient approach was proposed to maintain data appropriateness for data mining purposes even after anonymization [13, 25].

3.4. Inadequacy of current methods for heterogeneous documents

Sometimes, elimination of all identifiers is not even enough to preserve privacy in a special type of document [27]. Studies have shown that different identifiers and documents need to different anonymization approaches. For example, Omran et al. [27] dealt with a key problem through a known anonymization method named k-anonymity. Through k-anonymity, it is not possible to precisely determine which identifiers to be generalized and which to be suppressed. To solve this problem, an ontology-based strategy has been proposed. Moreover, the majority of anonymization methods have been evaluated on a special type of clinical data. Ferrández also indicated that an anonymization method developed for a specific document corpus cannot be appropriate for other types [28]. Among included publications, three papers focused on this issue [20, 27-28] (Table 2).

Table 2: Summary of the research results

Author	category	Aim	Related findings
Neubauer [18]	SU	Assessing existing privacy enhancing methods including anonymization, encryption, depersonalization, role-based access control and pseudonymization.	Pseudonymization supports the privacy of patients and keeps data accuracy intact for secondary usage.
Grechanik [19]	SU	Introducing a new view with which organizations can determine how much test coverage they can lose when using data privacy to database-centric applications.	Using *k*–anonymity (a data privacy approach) leads to serious degradation of test coverage.
Qingming [16]	SU, IE	A utility-based k-anonymity is proposed.	The proposed algorithm has completely less normalized certainty penalty (NCP) cost and it has lower query ratio than other algorithms. [+]
Harada [15]	SU	A new k-anonymity approach in which generalization hierarchies are automatically made by input information.	Automatically establishing generalization hierarchies decreases information loss following k-anonymization [*]
Loukides [17]	SU, IE	A new anonymization algorithm is suggested, which uses generalization and suppression to choose items, based on data publishers' utility needs, and is leaded by introducing utility criterion measure.	Using an efficient utility measure, this algorithm can be very useful at preserving data utility. It also allows more accurate query answering than other methods.

Author	category	Aim	Related findings
Benitez [22]	RR	Some techniques are introduced to estimate re-identification risk for many de-identification(De-ID) data sharing policies.	The differences in distributing population of U.S. states and their policies for disseminating datasets lead to varying re-identification risks.
Rothstein [24]	RR	Adverse effects of nonconsensual use of anonymized health data in research are included	De-ID was considered an important but insufficient means of keeping health privacy. It is indefensible from technical, ethical, and policy views to go on drawing a regulatory distinction between identifiable and de-identified health data.
Gellman [23]	RR	Risks and dangers to subjects and the research community are highlighted from use of supposedly anonymized information.	There is no ready enforcement for De-ID failures. The use of anonymized information for research goals should be regulated or even forbidden. Although, additional restrictions will make research impossible.
Khokhar [20]	SU, RR, IN	A cost-benefit evaluation is done to test related cost factors associated with the value of anonymized data and the possible damage cost due to privacy breaches.	The analytical cost model is efficient for health information custodians (HICs) to decide better on sharing health data for secondary and commercial uses.
Panackal [13]	IE	Introducing an adaptive utility based anonymization for accessing privacy without compromising the content of data or data mining accuracy	Both original and anonymized data sets are tested for classification accuracy and the conclusions showed that the anonymization process does not provide any important degradation in the accuracy of data mining classification.
Pruski [25]	IE	Introducing an ontology-based approach for efficient information retrieval in encrypted EHRs	The combined use of metadata and ontologies offers exciting features to improve, in terms of relevance, the results of a search. In addition, the uses of standard vocabularies make the construction and interpretation of the queries easier.
Deleger [26]	IE	(1) Evaluating the natural language processing (NLP)-based method to de-identify a large set of diverse clinical notes automatically. (2) Measuring the effect of De-ID on the performance of IE algorithms on the anonymized documents.	The performance of the system was indistinguishable from that of human annotators. The impact of automated De-ID was minimal on the utility of the narrative notes for subsequent IE as measured by the sensitivity and precision of medication name extraction.
Wu [12]	IE	To investigate the issue of health data utility after three anonymization methods with new criterions to assess the data utility (Support Vector Machine (SVM) and Earth Mover's Distance (EMD)$^{\times}$)	The results revealed that there is a significant difference in classification accuracy between evaluations on the original and anonymized data. In EMD experiment, it is shown that privacy preservation methods can significantly jeopardize the data utility due to the highly strict protection principles they impose.
Meystre [9]	SU , IE	The effect of five different De-ID methods was investigated based on clinical text information content (informative and formatting) and clinical information extraction by comparing counts of SNOMED-CT concepts found in the original and anonymized corpus.	The informativeness was only minimally altered by these systems while formatting was only changed by one system. Only about 1.2–3% less SNOMED-CT concepts were identified in anonymized corpus.

Author	category	Aim	Related findings
Gkoulalas-Divanis [2]	IE	Introducing a novel anonymization methods which is able to anonymize data with a desired balance between utility and privacy[α]	The results showed the relative error in query answering.
Liu [10]	IE	Evaluating performance of four De-ID methods that may be used to ensure regulatory compliance while also making practical database updating and querying easier.	Different De-ID methods have different effects on database operations such as time needed for data insertion, initial data loading and query on the database. Overall, De-ID has an undesirable effect on longitudinal study prevention.
Omran [27]	IN	A new ontology-base k-anonymization is proposed to determine which information can be generalized and which information needs to be suppressed.	The method could play an important role in protecting the privacy of personal health records without sacrificing the value of information for primary and secondary usages.
Ferrández [28]	IN	Various De-ID methods are comparing according to the generalizability and portability on different document sources as train and test sets.	There is no good report and results for these three systems as generalizability experiment.

[+] *Normalized certainty penalty (NCP) and Query answerability are two metrics that measure the utility of the data.*
[*] *Information loss is measured in terms of information entropy using a frequency distribution.*
[×] *Classification by SVM and evaluating the similarity between anonymized and original tables based on EDM are two approaches for investigating the utility loss of privacy preservation techniques.*
[α] *The utility policy constructed by Utility Policy Extraction (UPE) leads to the production of anonymized data that allows accurately computing the number of patients with the selected diseases.*

4. Discussion

The present study explored the efficiency of and the issues related to the anonymization of EHRs. It revealed that anonymization, though appropriate for preserving patients' privacy of healthcare data, cannot dispose of data re-identification risk altogether. On the other hand, when through the anonymization process a great portion of identifiable data is removed, the data will not be appropriate for a secondary use. This issue has been pinpointed in Meystre's study [29] under the title of over-scrubbing. Many anonymization methods have been suggested which mainly addressed data utility after anonymization. Such methods were explored in the first category i.e. data secondary use. The findings of the second category showed that if standard methods are followed for anonymization, there will be a lower risk of re-identification. Moreover, an accurate and detailed analysis of different types of re-identification risks and their effects could be helpful to data disclosure policy making and the right implementation of anonymization methods.

The main goal of data aggregation is analysis and use of the extracted information. The effect of anonymization on IE and database operations has been explored separately in the third category. Recent investigations proved that new anonymization methods have inconsiderable impacts on database operations, IE-based applications and text mining. Another issue taken into account was the text type. Unfortunately, the majority of anonymization methods have been evaluated on a specific type of clinical notes. Furthermore, they are mostly focused on English-language texts. However, the body of research selected in the fourth category showed that an anonymization method, once

designed for a certain type of clinical texts, will not produce desirable results on other text types.

All issues, covered here, addressed the significant issues related to anonymization domain which has got to be resolved through efficient approaches. However, this does not imply that anonymization is improper for privacy preservation.

The present study systematically reviewed the recent published researches about patient information anonymization. The effectiveness of this anonymization procedure was investigated in four categories: secondary use of anonymized data, re-identification risk, effect of anonymization on IE, and inadequacy of current methods for different text types. Although anonymization does not reduce the risk of re-identification to zero, if implemented correctly, could be useful in preserving patients' privacy. Moreover, a comprehensive analysis of different types of re-identification attacks plays a key role in data exposure policy making and developing anonymization algorithms.

References

[1] I. Anciu, J.D. Cowan, M. Basford, X. Wang, A. Saip, S. Osgood, et al, Secondary use of clinical data: the Vanderbilt approach, *J Biomed Inform* **52** (2014), 28-35.

[2] A. Gkoulalas-Divanis, G. Loukides, J. Sun, Toward smarter healthcare: Anonymizing medical data to support research studies, *IBM J Res Dev* **58**(1) (2014), 9:11.

[3] E.l. Emam, E. Jonker, L. Arbuckle, B. Malin, A systematic review of re-identification attacks on health data. *PloS one* **6**(12) (2011), e28071.

[4] P.C Kaur, T. Ghorpade, V, Mane, Analysis of data security by using anonymization techniques. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence); 14-15 Jan (2016).

[5] National Partnership foe women and Families, Making IT meaningful: How consumers value and trust health IT, Washington, DC, USA (2012).

[6] EU General Data Protection Regulation, https://www.eugdpr.org, last access: 3.3.2018.

[7] K.A. Wager, F.W. Lee, J.P. Glaser, *Health Care Information Regulations, Laws, and Standards*, in: Health Care Information Systems: A Practical Approach for Health Care Management. Wiley, San Francisco, CA., 2013. pp. 85-90.

[8] E. Khorshidi, Patient's Bill of Rights, http://www.e-khorshidi-lawyer.ir/index.php?ToDo=ShowArticles&AID=13145, last access: 3.3.2018.

[9] S.M. Meystre, O. Ferrandez, F.J. Friedlin, B.R. South, S. Shen, M.H. Samore. Text de-identification for privacy protection: a study of its impact on clinical text information content, *J Biomed Inform* **59** (2014), 142-150.

[10] J. Liu, S. Erdal, S.A. Silvey, J. Ding, J.D. Riedel, C.B. Marsh, et al, Toward a fully de-identified biomedical information warehouse. AMIA Annu Symp proc AMIA Symposium, (2009), 370-374.

[11] ISO, Health informatics – Pseudonymization, (2017).

[12] L. Wu, H. He, O.R. Zaïane, Utility of privacy preservation for health data publishing. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, (2013), 510-511.

[13] J.J Panackal, A.S. Pillai, Adaptive Utility-based Anonymization Model: Performance Evaluation on Big Data Sets, *Procedia Comput Sci* **50** (2015), 347-352.

[14] B.S. Elger, J. Iavindrasana, L.L. Iacono, H. Müller, N. Roduit, P. Summers, et al. Strategies for health data exchange for secondary, cross-institutional clinical research, *Comput Methods Programs Biomed* **99**(3) 2010 ,230-251.

[15] K. Harada, Y. Sato, Y. Togashi, Reducing Amount of Information Loss in k-Anonymization for Secondary Use of Collected Personal Information. 2012 Annual SRII Global Conference, (2012).

[16] T. Qingming, W. Yinjie, L. Shangbin, W. Xiaodong, Utility-based k-anonymization, The 6th International Conference on Networked Computing and Advanced Information Management, (2010).

[17] G. Loukides, A. Gkoulalas-Divanis, Utility-preserving transaction data anonymization with low information loss, *Expert Syst Appl*, **39**(10) 2012, 9764-9777.

[18] T. Neubauer, B. Riedl, Improving patients privacy with Pseudonymization, *Stud health technol inform* **136** (2008), 691-696.

[19] M. Grechanik, C. Csallner, C. Fu, Q. Xie, Is Data Privacy Always Good for Software Testing? 2010 IEEE 21st International Symposium on Software Reliability Engineering, (2010).

[20] R.H. Khokhar, R. Chen, B.C. Fung, S.M. Lui, Quantifying the costs and benefits of privacy-preserving health data publishing, *J Biomed Inform*, **50** (2014), 107-121.
[21] H. Kikuchi, T. Yamaguchi, K. Hamada, Y. Yamaoka, H. Oguri, J. Sakuma, Ice and Fire: Quantifying the Risk of Re-identification and Utility in Data Anonymization, 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), (2016).
[22] K. Benitez, B. Malin, Evaluating re-identification risks with respect to the HIPAA privacy rule, *J Am Med Inform Assoc*, **17**(2) 2010, 169-177.
[23] R. Gellman, Why deidentification fails research subjects and researchers, *Am J Bioeth* **10**(9) (2010), 28-30.
[24] M.A. Rothstein, Is deidentification sufficient to protect health privacy in research?, *Am J Bioeth* **10**(9) (2010), 3-11.
[25] C. Pruski, F. Wisniewski, Efficient medical information retrieval in encrypted Electronic Health Records, *Stud Health Technol Inform* **180** (2012), 225-229.
[26] L. Deleger, K. Molnar, G. Savova, F. Xia, T. Lingren, Q. Li, et al. Large-scale evaluation of automated clinical note de-identification and its impact on information extraction, *J Am Med Inform Assoc*, **20**(1) (2013), 84-94.
[27] E. Omran, A. Bokma, S. Abu-Almaati, A K-anonymity Based Semantic Model For Protecting Personal Information and Privacy, IEEE International Advance Computing Conference, (2009).
[28] O. Ferrandez, B.R. South, S. Shen, F.J. Friedlin, M.H. Samore, S.M. Meystre, Generalizability and comparison of automatic clinical text de-identification methods and resources, AMIA Ann Symp proc, (2012), 199-208.
[29] S.M. Meystre, F.J. Friedlin, B.R. South, S. Shen, M.H. Samore, Automatic de-identification of textual documents in the electronic health record: a review of recent research, *BMC Med Res Methodol* **10** (2010).

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-88

Exporting Data from a Clinical Data Warehouse

Georg FETTE[a,b,1], Mathias KASPAR[b], Leon LIMAN[a], Georg DIETRICH[a], Maximilian ERTL[b], Jonathan KREBS[a], Stefan STÖRK[b], Frank PUPPE[a]
[a] *University of Würzburg, Chair of Computer Science 6*
[b] *University Hospital of Würzburg, Comprehensive Heart Failure Center*

Abstract. Data Warehouses (DW) are useful tools to support clinical studies as they can provide exports of routine care data for scientific reuse. Exported DW data is usually post-processed and integrated into study databases by study staff that is reasonably trained in specific tools like SPSS and Excel but which are no programmers or computer scientists. DW systems should therefore be configurable to satisfy export format desiderata as much as possible so that exports contain no unnecessary post-processing obstacles. In the presented work the authors analyze various existing DW systems in respect to a list of potential export formats.

Keywords. Clinical data warehouse, export data format

1. Introduction

An important element of clinical studies is the collection of patient data in so-called Case Report Forms (CRF). Routine clinical data contained in Data Warehouses (DW) can be a valuable resource for automatic [1] or manual completion of CRFs. Often the transfer of DW data into a study database incorporates (semi-)manual post-processing steps using tools like SPSS, R or Excel. In many cases, data exported from the DW has to suffice format constraints, i.e. preferably a table with one patient per row and one column per attribute type (e.g. age, sex, laboratory value, etc.) to facilitate calculations of statistics with the mentioned tools. Although DWs should be designed to be generic tools capable for all sort of data exchange, export APIs of existing solutions have limited capabilities. Especially in cases with multiple measurements of desired attributes for a single patient, an easy-to-use, all-satisfying data export can pose difficulties.

In the presented work, the authors analyze various existing DW systems in respect to a list of potential export formats. In this matter, the focus lies on structural formats (e.g. table, tree- or graph-structure) rather than technical format specification (e.g. JSON, XML) as every structural format can be encoded in each technical format. The technical data format in which data is stored in respective DW systems (e.g. relational database, Entity-Attribute-Value-schema (EAV) [2]) is as well neglibable, as the structural relationships between returned data elements remain the same.

[1] Corresponding author: Georg Fette, University Hospital of Würzburg, DZHI, Am Schwarzenberg 15, 97078 Würzburg, Germany; E-Mail: georg.fette@uni-wuerzburg.de

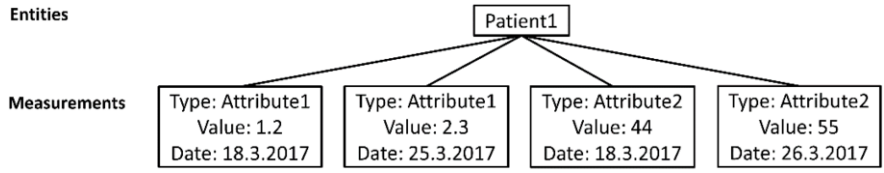


Figure 1. Data example with one patient having four different measurements.

2. Methods

The challenges arising when exporting data from a DW shall be exemplified with the following small piece of data that is assumed to be found when querying for two different attributes. The query results in one patient that has two measurements measured at two different dates for each of the two attributes. The returned data is depicted in Figure 1.

In the following examples the measurement dates are omitted. If those are desired as well, one additional date column would have to be added next to each value column. Multiple possible format solutions exist to export the data depicted in Figure 1:

1) Each attribute gets assigned a reduction function that deterministically reduces multiple measurements of one attribute to one measurement (Table 1).
2) Multiple measurements for one attribute are concatenated, separated by a specific character and written to the same result table cell (Table 2).
3) Multiple measurements are distributed over multiple rows. Multiple variants are possible for this solution:
   a) Different attributes each get their own column. The rows are a Cartesian product of all measurements of all attributes (Table 3).
   b) Different attributes each get their own column. Each row contains only values of exactly one attribute (Table 4)
   c) The values of different attributes are written to the same column. An additional column indicated the attribute type that belongs to each row (Table 5). This format is commonly known as an EAV-schema [2].
4) For each attribute the result table contains as many columns equal to the maximum number of measurements for that attribute for one patient. (Table 6)
5) The data is not exported as a table but instead in a tree-structure form.

Table 1. Result data using the reduction function "minimum(x1, x2, ...)"

PatientID	Attribute1_min	Attribute2_min
Patient1	1.2	44

Table 2. Result data with measurements of the same attribute being concatenated and written to the same cell

PatientID	Attribute1	Attribute2
Patient1	1.2; 2.3	44; 55

Table 3. Result data with measurements of the same patient being distributed over multiple rows. The table is the Cartesian product of all measurements of all attributes.

PatientID	Attribute1	Attribute2
Patient1	1.2	44
Patient1	2.3	44
Patient1	1.2	55
Patient1	2.3	55

Table 4. Result data with measurements of the same patient being distributed over multiple rows with data from different attributes in separate sets of rows

PatientID	Attribute1	Attribute2
Patient1	1.2	
Patient1	2.3	
Patient1		44
Patient1		55

Table 5. Result data formatted in an EAV schema

PatientID	AttributeName	Value
Patient1	Attribute1	1.2
Patient1	Attribute1	2.3
Patient1	Attribute2	44
Patient1	Attribute2	55

Table 6. Result data with multiple columns for each attribute with multiple measurements

PatientID	Attribute1-1	Attribute1-2	Attribute2-1	Attribute2-2
Patient1	1.2	2.3	44	55

Existing DW systems were analyzed in their export capabilities in respect to the listed export format solutions. The systems analyzed were the DW system i2b2 [3], an OMOP CMD data model queried with SQL [4], Medical Research Insights (MRI) from SAP [5], the openEHR query API [6] and the DW system PaDaWaN [7].

3. Results

i2b2 is an openly available DW framework. In its version 1.7.09c the web client contains a data export plugin that allows to export selected attributes belonging to a patient set of a recently executed query. The plugin allows the selection of either solution 2, 3b or 3c. The selected solution is applied to the complete set of desired attributes to be exported.

OMOP CDM is not a DW system but a relational database data model that can be queried with SQL. The constraints using SQL as query method do not only hold for OMOP CDM, but for every relational data model queried with SQL. Therefore, OMOP CMD stands as a representative for DW systems without a dedicated query and export engine, in which users instead have to write SQL queries in order to access data. The main mechanism in SQL for retrieving data of different attributes and combining them to data sets belonging to individual patients is the JOIN operator. The straight-forward application of JOINs creates tables represented by solution 3a. When the query process is instead modeled in a way that first a set of matching patient-IDs are selected and then desired attributes facts belonging to those IDs using the UNION operator are collected, result tables represented by solution 3c can be realized. As SQL is a Turing-complete language, every one of the proposed solutions could be realized, although this would be at the cost of a much more complex implementation effort.

MRI is a commercial clinical DW developed by SAP. It is still in ongoing development and in its current version 2.0 it only provides a data export format represented by solution 2.

openEHR is a freely available DW specification for which free implementations as well as commercial solutions exist. The currently proposed API 0.9 (which is still in development) does not explicitly define which solutions are supported, but its current

documentation and the discussions about it in the openEHR mailing list archives lets suspect solution 3a or a variant of it (see [6] *ResultSet example*).

PaDaWaN is a DW framework developed at the University Clinic of Würzburg. In PaDaWaN individual attributes of a query can be parametrized independently to use one of the solutions 1, 2 or 3a. Solution 3a is only allowed for one of the attribute in a query to prevent potential exponential explosions concerning the amount of returned rows.

For i2b2 and PaDaWaN any solution could be provided with additional manual work using R plugins that provide additional formatting options [1, 8].

4. Discussion

The selection of the presented solutions was chosen based on existing solutions in the analyzed DW systems as well as due to feature requests brought to the authors by local study staff. Solutions 4 and 5 are such feature requests that are not provided by any of the analyzed systems, but still have been included in the list of possible solutions.

We first want to discuss strengths and weaknesses of the presented solutions: Solution 1, 2 and 4 have the advantage that the result table contains as many rows as patients. This can be useful when a user uses simple viewer tools like Notepad or Excel to quickly check how many patients have been found or when an export has to be joined with a new one into a single document.

Solution 1 is an easy to use solution that prevents unwanted bloating of the result table concerning the amount of rows as well as columns. Unfortunately, a data-reduction takes place when executing the given reduction function during the export. When study staff processes data exported in that format this always has to be kept in mind.

Solution 2 is potentially well suited for manual screening of the exported data. Automatic post-processing of data exported in this format is difficult as the concatenated lists have to be split again before any further processing. Sorting of the result table using tools like Excel is hardly possible as formerly separate values are now concatenated texts.

All variants of Solution 3 have the disadvantage that the result table can contain more rows than patients. This hinders a quick estimation of patient counts when manually analyzing documents in this format. Furthermore, the result tables cannot be easily resorted as unskillful sorting by one attribute can destroy the arrangement of row sets originally belonging to the same patient.

Solution 3a creates, due to its Cartesian product nature, the risk of an exponential explosion regarding the amount of created rows depending on the amount of attributes and the amount of measurements for each attribute. Having a patient with n attributes with m measurements each would create m^n result rows. This can be a problem for queries with many attribute parameters having many measurements.

Solution 3b as well as 3c are reasonably well human as well as machine readable. A property of solution 3b (that is not really a disadvantage but still noteworthy) is that the result tables are quite sparse because always only one attribute column is filled with data. Both solutions 3b and 3c create for queries with n attributes for one patient with m measurements per attribute n*m result rows.

Solution 4 is a format commonly used in the design of study CRFs. A disadvantage of the format is that two exports using the same query can create differently sized output tables. This can happen when the query returns different amounts of measurements for the same attribute, depending on the data contained in the database on which the query is carried out. Furthermore, the amount of created columns can grow quite large, which

can pose problems when importing the result table in tools having a limited maximum number of allowed columns (e.g. 16,384 in Excel). Another problematic disadvantage of solution 4 emerges when performing post-processing steps on data exported in such a format: All functions that have to incorporate all measurements of one attribute have to be parametrized with all columns containing values of that attribute. All such columns have to be included in large Or-clauses (e.g. "*IF (attribute1 > 2)*" → "*IF (attribute1-1 > 2 OR attribute_1-2 > 2)*"). Always having to list all columns of an attribute leads to bloated post-processing scripts and is error prone as possibly columns can be incorrectly forgotten in the needed listings.

Solution 5 is a format that is represented by data interchange formats like CDISC ODM or FHIR. CDISC ODM is a widely accepted standard for the exchange of clinical data and there exist approaches for the import of CDISC ODM data into data processing tools like SPSS [9]. Unfortunately, many data clerks are not very familiar with the handling of tree-structure based formats using common data processing tools, like SPSS or Excel. In Excel, the import of tree-structure based formats is not possible at all.

As all solutions have advantages as well as disadvantages it would be preferable to switch between solutions as easy as possible. Furthermore, a user should have the possibility to parametrize an export so that the potential reusability of that export is maximized. Therefore, a parametrization that allows to choose each of the presented solutions fine-grained for each individual query attribute would be beneficial, which is currently only possible in PaDaWaN. Solution 5 is, due to its strongly different format nature, hardly combinable with the other solutions.

When assessing the different analyzed DW systems concerning the presented solutions PaDaWaN supports the most solutions in the most fine-grained manner (i.e. per attribute). i2b2 supports at least three of the presented format solutions although the selected choice is always affecting the format of all attributes of the respective query. In all other DW solutions, users have to cope with only one supported solution.

As all analyzed systems are still in active development or are even open source projects, the presented work could be an incentive to improve the respective data export capabilities or to develop adapters that provide additional functionalities.

Acknowledgement

This research was funded by grant of German Federal Ministry of Education and Research (Comprehensive Heart Failure Center Würzburg, grants #01EO1004 and #01EO1504).

References

[1] Kaspar M, Ertl M, Fette G et al, Data linkage from clinical to study databases via an R data warehouse user Interface. Experiences from a large clinical follow-up study, Methods Inf Med. 55 (2016), 381–386.
[2] Dinu V, Nadkarni P, Guidelines for the effective use of entity-attribute-value modeling for biomedical databases, Int J Med Inform 76 (2007), 769-779.
[3] Murphy S, Weber G, Mendis M et al, Serving the Enterprise and beyond with Informatics for Integrating Biology and the Bedside (i2b2), J Am Med Inform Assoc 17 (2010), 124-130.
[4] OMOP CDM Queries, http://cdmqueries.omop.org/, accessed 2.3.2018
[5] SAP Medical Research Insights (MRI), https://icn.sap.com/projects/sap-medical-research-insights.html, accessed 2.3.2018

[6] OpenEHR REST Query API (v0.9 proposal), https://www.openehr.org/releases/ITS/Release-0.9.0/docs/query.html#header-resultset-example, accessed 2.3.2018
[7] Fette G, Dietrich G, Ertl M et al, Die grafische Benutzeroberfläche PaDaWaN für das klinische Data Warehouse für Routinedaten an der Universitätsklinik Würzburg, Proceedings GMDS 73 (2015), 152.
[8] Segagni D, Ferrazzi F, Larizza C et al, R engine cell: integrating R into the i2b2 software infrastructure, J Am Med Inform Assoc. 18 (2011), 314-317.
[9] Bruland P, Dugas M, S2O - A software tool for integrating research data from general purpose statistic software into electronic data capture systems, BMC Med Inform Decis Mak 17 (2017).

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-94

A Clinical Data Warehouse Based on OMOP and i2b2 for Austrian Health Claims Data

Christoph RINNER[a,1], Deniz GEZGIN[a], Christopher WENDL[a] and Walter GALL[a]
[a]*Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Austria*

Abstract. *Background:* To develop simulation models for healthcare related questions clinical data can be reused. *Objectives:* Develop a clinical data warehouse to harmonize different data sources in a standardized manner and get a reproducible interface for clinical data reuse. *Methods:* The Kimball life cycle for the development of data warehouse was used. The development is split into the technical, the data and the business intelligence pathway. *Results:* Sample data was persisted in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The i2b2 clinical data warehouse tools were used to query the OMOP CDM by applying the new i2b2 multi-fact table feature. *Conclusion:* A clinical data warehouse was set up and sample data, data dimensions and ontologies for Austrian health claims data were created. The ability of the standardized data access layer to create and apply simulation models will be evaluated next.

Keywords. Secondary use, standardized health data, clinical data warehousing.

1. Introduction

Clinical data reuse is defined as "non-direct care use of personal health information including but not limited to research" [1]. In [2] it is concluded that clinical data reuse can constitute an important pillar to increase the quality in medical research and efficiency of clinical research. Reused data can come from various sources, documented for various purposes in varying granularities and levels of structure. To successfully reuse these data, the source data have to be cleaned, harmonized and pre-processed to allow meaningful research.

Various efforts exist to unify clinical data reuse by storing data in harmonized data models. The Observational Health Data Sciences and Informatics (OHDSI) program [3] supports the community in the development and adaptation of the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). Informatics for Integrating Biology and the Bedside (i2b2) [4] offers a generic data model for clinical data and an open source clinical data warehouse framework with advanced access and querying mechanisms and good extensibility using so called cells. In [5] the CDISC Operational Data Model (ODM) was implemented using i2b2 as front-end.

A clinical data warehouse offers life science businesses easier access to stored information and gives insights into their health care data. As part of the imProve project[2]

[1] Corresponding Author: Christoph Rinner, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria, E-Mail: christoph.rinner@meduniwien.ac.at.

[2] http://www.dwh.at/de/expertise/projekte/improve/

we want to use simulations of the Austrian health care landscape to innovate the health care industry. Simulation models are models of reality which are created and parameterized by the model builders using information resources describing reality. The expected benefit of a clinical data warehouse with a standardized interface between the developed simulation models and the source data is easier deployment of the models and reproducible results.

This article describes the process of implementing a local clinical data warehouse infrastructure based on the OMOP CDM and the i2b2 data warehouse infrastructure as front-end to analyze Austrian health claims data.

2. Methods

Based on the Kimball Lifecycle diagram [6] a local clinical data warehouse infrastructure was created. The business requirements are specified first, and then the lifecycle splits up into three pathways. In the technical pathway strategic and technical directions are planned based on the business requirements. The data pathway reflects the categorization of data into measurement facts and descriptive dimensions. Dimensions can be seen as structures that categorize the measurement facts to enable users to answer business questions. First the relevant dimensions based on the available data (i.e. measurement facts) are analyzed and formalized, then the data models to store the measurement facts are created and finally the source data are loaded into the data warehouse. In the business intelligence pathway the information needs are analyzed and tools to access and query the data are created.

3. Results

3.1. Business Requirements

As a first use case Austrian health claims data were selected. Health claims data are documented in a highly structured form needed for reimbursement purposes and existing know-how from previous projects with this type of data was available. To prevent data privacy concerns when using real claims data sample data was created. The clinical data warehouse should be extendible for future use cases in respect to the clinical data stored but also in respect to technical interfaces like future distributed query possibilities. The data stored in the data warehouse should be accessible during the building of simulations models to parameterize the model and as input during the execution of simulations models.

3.2. Technical path

As part of a previous project [7], an i2b2 clinical data warehouse was deployed for Austrian health claims data. A special focus was put on technical aspects like the handling of large data sources and compartmentalization using Docker containers. I2b2 implements various strategies to maintain patient privacy [8] and offers a generic database design based on the star schema. Observations and the clinical data are stored in five main tables. This generic structure allows for a high degree of flexibility yet

increases complexity. In the current version of i2b2 the multi-fact table concept is introduced to enable i2b2 to also access data from other data sources beside the i2b2 observation fact table[1]. This feature allows for example OMOP CDM as source for observational data in i2b2 using database entries without changes to the i2b2 source code. OMOP CDM is optimized for claims data with separate tables for drug costs, visits costs etc. resulting in an easier to understand column-based database design.

Based on the business requirements the OMOP CDM (v5.0, October 2014) and i2b2 (v1.7.09b) accessing OMOP CDM via the new multi-fact table feature was selected. For the deployment a virtual machine based on VMWare ESXI (v5.5), with 1 CPU (3 GHz), 12 GB RAM and 130 GB of disk space was created. As operating system Ubuntu server (v16.04.2) was selected to host i2b2 and the Microsoft SQL Server (v14.0).

3.3. Data path: Analysis of data dimensions

The OHDSI Athena[2] allows the distribution of standardized vocabularies, which already use the correct data format for the OMOP CDM. The two dimensions for the Anatomical Therapeutic Chemical (ATC) Classification System used to document dispensed drugs and the International Statistical Classification of Diseases and Related Health Problems 10th revision (ICD-10) used to document hospital diagnoses in Austrian health claims data could be reused directly.

Additionally to the ATC codes, pharmaceutical products sold in Austria are also identified using the unique Austrian pharmaceutical registration number (PRN). The 16,500 medications represented by unique PRN are modeled as separate *concepts* (with distinct entries for *vocabulary* and *concept_class*) in the OMOP CDM. Using *concept_relationship* the PRN are linked to the existing ATC codes.

The catalogue of 1,951 individual medical services (i.e. “MEL codes”) was hierarchically structured into chapter, subchapter, unit, anatomy (coarse and fine), access and source. Each MEL code and structure was modeled as new *concepts* and linked using the *concept_relationships*.

Austria can be divided into more than 17.000 cities (i.e. “Orte”) with distinct postal codes located in 2,100 counties (i.e. “Gemeinden”) located in in 32 supply regions (i.e. “Versorgungsregionen”) located in 9 states (i.e. “Bundesländer”) and finally four supply zones (i.e. “Versorgungszonen”). The *location* table in OMOP CDM has fields for all entities except the supply zones and regions. Therefore these two dimensions were documented in the *address_2* field instead of creating separate *observations* for each visit.

Claims data in Austria originates from the 19 insurance carriers which were modelled as *concepts*. For each *visit_occurence* one of 281 *care_sites* was assigned including the unique hospital id, state, supply region and supply zone.

3.4. Data path: Physical design of OMOP Schema

From the 36 tables available in the OMOP CDM only 15 were used to document the Austrian health claims data. The *person, death* and *location* tables are used to store information about the patients. The *provider* table (including *care_site* and *location*) are used to store information about the health care provider. Hospital visits were modelled as *visit_occurences* with *observations*, *measurements* and *condition_occurence* to

[1] https://community.i2b2.org/wiki/display/MFT/Multi-fact+Table+Home
[2] https://www.ohdsi.org/analytic-tools/athena-standardized-vocabularies/

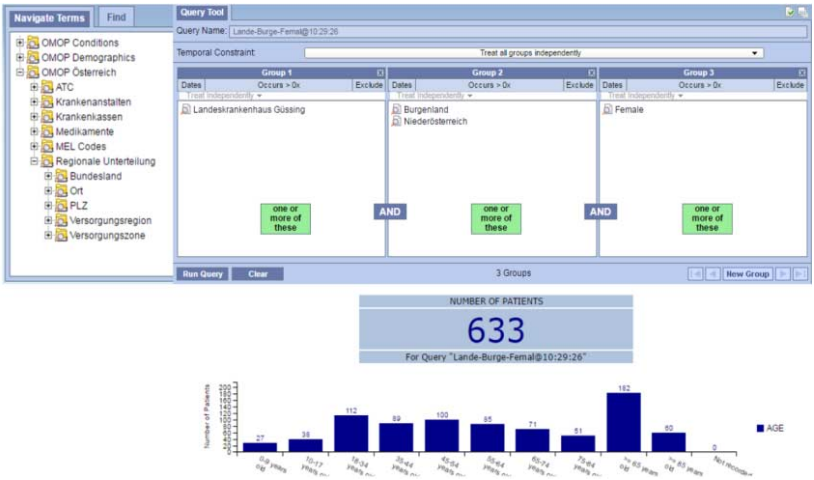


Figure 1: Overview of the i2b2 query editor with the created dimension on the left hand side, a sample query in the center and the results at the bottom.

document information about the visit. Prescribed and dispensed medications are stored in the *drug_exposure* table. The previously described data dimensions are modelled using the tables *concept*, *concept_class*, *domain*, *vocabulary* and *concept_relationship*.

3.5. Data path: Integration of the claims data

To evaluate the created clinical data warehouse infrastructure and to prevent data privacy issues we created sample data for female Austrian patients base on the distribution of age, discharge diagnoses and medication distribution in the Austrian population. The final sample data consists of 4.3 million patients with 460.000 hospital stays and 10.3 million medications representing 16.500 different pharmaceutical products and covered a time period of 3 years. We initially evaluated the Observational Medical Dataset Simulator (OSIM)[1] for OMOP. Due to version incompatibility of OSIM we could not apply it directly and decided to created comma separated value (csv) files using a simple JAVA routine with the distributions as input. All csv files were imported into the database and all transformations were performed using SQL scripts. This approach, where the data wrangling and transformations are performed directly in the database instead of separate tools is called extract, load and transform (ELT) process [9].

3.6. Business Intelligence

To get a generic overview of the imported data in the OMOP CDM the Achilles[2] framework was used. To allow more in-depth analysis and to control access on a per user base, the i2b2 web interface was used. In Figure 1 an overview of the query interface with the various dimensions is depicted. To access OMOP CDM with i2b2 the i2b2 multi-fact tables were enabled and the OMOP CDM data was linked to i2b2 by creating a dedicated i2b2 ontology for our data. This OMOP_AUSTRIA ontology references the OMOP CDM tables directly. For complex queries (e.g. “Patients with more than 3

[1] ftp://ftp.ohdsi.org/osim2/
[2] https://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/

hospital stays") we created separate database views also accessible using the multi-fact feature. The final goal will be to create simulation models based on the data stored in the data warehouse and offer access to the simulations models to the end users.

4. Discussion and Conclusion

We implemented a clinical data warehouse based on i2b2 using the new i2b2 multi-fact tables to access health claims data stored in the OMOP CDM. Austrian dimensions for pharmaceutical registration numbers, regions, health care providers, insurance carriers and medical services were created. Sample data of more than 4 million patients was created and imported using SQL scripts. Using generated data and health claims data in particular, the import was quite straight forward since data were already structured and no data cleaning, data extraction or linking was needed. The first cycle of the Kimball life cycle for data warehouses was finished.

The created data warehouse can be deployed at different data holders using the virtual image created. To deploy the data warehouse extract, load and transformation steps are still needed to import the source data. To create simulation models, the Achilles tool is a good starting point to gain an overview of the stored data and for quality assessment. Depending on the use case specific reports (e.g. regional distribution) should be developed. The i2b2 interface allows direct access to cohort estimations on a very fine grained level (e.g. number of patients with a specific diagnosis in a specific time frame followed by a specific medication). If the data holders are not willing to offer direct access to their data due to data privacy regulations and privacy concerns, Achilles and i2b2 enable the data holder to extract needed parameters themselves without advanced technical background.

As a next step we plan to quality assess the developed dimensions and distribute them as standardized vocabularies with the Athena tool from the OHDSI community. In a similar fashion the i2b2 ontology we created to query OMOP CDM data representing Austrian health claims data with i2b2 should be made available to other researchers. We plan to refine the import scripts to automate the import process further and to allow all Austrian health care providers documenting health claims data or insurance carriers receiving health claims data to easily reuse the created dimension and store their data in their locally deployed OMOP CDM. This is a first step to prepare a small subset of dimensions used in the Austrian health care landscape for reuse purposes.

The OMOP CDM and the i2b2 ontology offer a good framework to formalize dimensions. Reusing these dimensions can help to facilitate future data preparation tasks and make research easier to compare. At the time of the first deployment, the multi-fact tables in i2b2 were only optimized for Microsoft SQL servers. Since these are now also available for PostgreSQL we are currently migrating the dimensions, ontologies, import scripts to PostgreSQL. The OMOP CDM allows an easier selection of stored data using SQL, compared to the generic i2b2 data model. I2b2 only distinguishes between patients, visits, observations and observers, all other nuances in the data are hidden behind the concept dimension. OMOP offers the Achilles tools to get a quick overview of the data in the data warehouse and with the i2b2 query editor a simple interface for end users is available.

When analyzing claims data OMOP CDM already offers specific tables for medications services. With harmonized ontologies in i2b2 the same effect can be achieved. Combining the data model of OMOP CDM and the query interface and data

management of i2b2 is relatively easy to achieve and enhances both products. Our proposed data warehouse implementation can build a starting point for other projects to enable reproducible and comparable research and facilitate the development and application of simulations models.

Acknowledgement

This research was supported by Vienna Business Agency. We thank all our collaborators in the imProve project especially Georg Duftschmid, Nicolas Popper and Barbara Glock.

References

[1] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, et al. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association : JAMIA. 2007;14(1):1-9.

[2] Meystre SM, Lovis C, Burkle T, Tognola G, Budrionis A, Lehmann CU. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. Yearbook of medical informatics. 2017;26(1):38-52.

[3] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Studies in health technology and informatics. 2015;216:574-8.

[4] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). Journal of the American Medical Informatics Association : JAMIA. 2010;17(2):124-30.

[5] Meineke FA, Staubert S, Lobe M, Winter A. A comprehensive clinical research database based on CDISC ODM and i2b2. Studies in health technology and informatics. 2014;205:1115-9.

[6] Kimball R. The data warehouse lifecycle toolkit: John Wiley & Sons; 2008.

[7] Endel F, Duftschmid G. Secondary Use of Claims Data from the Austrian Health Insurance System with i2b2: A Pilot Study. Studies in health technology and informatics. 2016;223:245-52.

[8] Murphy SN, Gainer V, Mendis M, Churchill S, Kohane I. Strategies for maintaining patient privacy in i2b2. Journal of the American Medical Informatics Association : JAMIA. 2011;18 Suppl 1:i103-8.

[9] Soutier M. Von ETL zu ELT in Big Data-Systemen 2015. Available from: http://www.soutier.de/blog/2015/03/01/von-etl-zu-elt-big-data/.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-100

EHR Text Categorization for Enhanced Patient-Based Document Navigation

Markus KREUZTHALER[a,d,1], Bastian PFEIFER[a], José Antonio VERA RAMOS[a], Diether KRAMER[c], Victor GROGGER[c], Sylvia BREDENFELDT[c], Markus PEDEVILLA[c], Peter KRISPER[b], and Stefan SCHULZ[a]
[a]*Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria*
[b]*Division of Nephrology and Dialysis, Department of Internal Medicine, Medical University of Graz, Austria*
[c]*KAGes Steiermärkische Krankenanstaltengesellschaft m.b.H., Graz, Austria*
[d]*CBmed GmbH - Center for Biomarker Research in Medicine, Graz, Austria*

Abstract. Patients with multiple disorders usually have long diagnosis lists, constitute by ICD-10 codes together with individual free-text descriptions. These text snippets are produced by overwriting standardized ICD-Code topics by the physicians at the point of care. They provide highly compact expert descriptions within a 50-character long text field frequently not assigned to a specific ICD-10 code. The high redundancy of these lists would benefit from content-based categorization within different hospital-based application scenarios. This work demonstrates how to accurately group diagnosis lists via a combination of natural language processing and hierarchical clustering with an overall F-measure value of 0.87. In addition, it compresses the initial diagnosis list up to 89%. The manuscript discusses pitfall and challenges as well as the potential of a large-scale approach for tackling this problem.

Keywords. Cluster Analysis, Natural Language Processing, Electronic Health Records, Semantics

1. Introduction

Patients with chronic diseases and multiple diagnoses accumulate high numbers of in- and outpatient treatment episodes. This leads to an overloaded content of their electronic health records (EHR). This fact especially hampers a quick visual perception of the most important patient-level information. In order to support physicians in accessing this information within a reasonable period of time, our goal is to provide them with a so-called QuickView, accessible from within the user interface of the hospital information systems (HIS). This enables the physician to get a navigable overview of the most important categories. One of these categories is a listing of all diagnosis phrases most of them coded by ICD-10. In case of multi-morbid patients this easily amounts to a listing of hundred or more diagnosis phrases. This list, which resembles the classical problem list, is the matter of investigation of our work. Due to its high redundancy, we intend to

[1] Corresponding Author: Markus Kreuzthaler, CBmed GmbH - Center for Biomarker Research in Medicine, Stiftingtalstrasse 5, 8010 Graz, Austria; Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerplatz 2/V, 8036 Graz, Austria, E-Mail: markus.kreuzthaler@medunigraz.at

apply a topic-based grouping, which can be exploited in a navigational and information visualization based implementation, settled within a web-based QuickView front-end.

The paper is organized as follows: Section 2 gives an overview of related work and highlights the need of our investigations. Section 3 describes the data sets, models the problem and presents the methods in use. The results are presented in Section 4 and a conclusion and outlook is given in Section 5.

2. Background and Related Work

In the last years, the analysis of large-scale EHR content using supervised and unsupervised machine learning has become a popular approach to gain more fine-grained insights into clinical information like diagnoses or medications. Unstructured EHR data like clinical narratives hamper efficient information extraction due to numerous language idiosyncrasies like short forms (abbreviations, acronyms), spelling and typing mistakes, syntactic incompleteness, specialist jargon, negations or non-standardized numeric expression, just to mentions some. A focus of interest has been the automatic code assignment (e.g. of ICD diagnosis codes) due to their relevance for billing, medical decision support and therapy planning.

Koopman et al. [1] introduced a method to automatically detect cancer diagnoses and classify them into ICD-10 codes. In a supervised framework, they used support vector machines (SVMs) with term and concept based features. They report an F-measure of 0.7 for detecting the type of cancer. In another work, Koopman et al. [2] also tried to automatically classify death certificates taking into account diabetes, influenza, pneumonia and HIV. Applying a supervised approach using SVMs, an F-measure of 0.8 was obtained for ICD-10 coding. Ning et al. [3] proposed an ICD-10 code assignment approach for Chinese medical text. They hierarchically structured the ICD-10 codes and assigned an unlabelled document based on a word-to-word similarity metric. An F-measure accuracy of 0.91 is reported. Chen et al. [4] improved the longest common subsequence algorithm for ICD-10 mapping to Chinese clinical narratives, yielding an F-measure of 0.81 for this task. Boytcheva [5] applied ICD-10 code mapping to Bulgarian clinical narratives. A multi-class SVM with a max-win voting strategy in combination with a text pre-processing module was applied for assigning ICD-10 codes, resulting in an F-measure of 0.84.

Most of the work mentioned here and a wide range of related studies deal with medical free text in a specific language, like we report on our data set written in German. Features used in a supervised framework as in the work cited above are often connected to language-dependent patterns, even though more recent approaches like deep learning methods reduce the need for use case specific feature engineering.

In the next sections we will introduce an approach without the need of feature engineering, which aims at a maximum of language-independence. We refrain from a purely supervised approach to achieve the goal of semantic grouping of patient based diagnosis list (problem list) items.

3. Methods and Data

3.1. Data Set

The initial experiments on content-based grouping include diagnosis listings of five de-identified nephrology patients, each of them having between 250 and 861 diagnosis statements written in German, accumulated within a time range of 12 to 22 years. A peculiar feature of these code-description pairs is the fact that physicians are allowed to overwrite the contents of a 50-character long standardized text field, filled with standard text produced by an ICD coding plugin. The resulting listing is therefore a mixture of standard and personalized diagnosis texts, the latter often enriched by adding procedures and dates. Text entries without ICD codes are also possible. As a result this diagnosis list comes close to what is well-known as ”problem lists” in the Anglo-Saxon medical tradition, but unknown in our context. The result is an admixture of a) list items in which the ICD description remains unaltered b) list items in which the ICD description is partly or fully overwritten and / or abridged c) list items with no ICD code attached. Especially in the case of patients with long clinical histories, the accumulation of list items lead to a high degree of redundancy.

3.2. Problem Modeling

This section gives a brief formal problem description and presents the underlying assumptions we make for diagnoses text grouping. Accordingly, every patient $P_{1..i}$ has a set of diagnosis list items $I_{1..k..l}$ where I_k = (ICD-10$_k$, $\mathbf{d}_k$) defines the coded ICD-10$_k$ 50-character long description $\mathbf{d}_k$ which we refer to as a document throughout the manuscript. One fraction $I_{coded} = I_{1..k}$ is coded and the other one $I_{non\text{-}coded} = I_{k+1..l}$ is without codes, with just the text snippets $\mathbf{d}_{k+1..l}$ existing. Since an immediate overview of all list items $I_{1..l}$ to a patient P_i is impossible with longer lists, our solution attempts to semantically group them into n sets $C_{1..n}$, so that the content navigation through all list items $I_{1..l}$ via $C_{1..n}$ is supported.

For semantically grouping of related list items $I_{1..l}$, we make the following assumptions: **Assumption A.** Disease list items I_{coded} with the same 3-digit ICD-10 code are similar in content by definition. This derives from our experience that assignments of the fourth digit tend to be ambiguous. If codes exist, they form a manual ground truth of judgement for semantic similarity. **Assumption B.** A content similarity of a sub group of list items $I_{i..j}$ out of $I_{1..l}$ is given by string similarity between two list items (I_1, I_2), which can be expressed via a function $f_{sim}(I_1, I_2) = sim = f_{sim}(\mathbf{d}_1, \mathbf{d}_2)$. Therefore *sim* is an indicator for content similarity.

The problem description and the *Assumptions A and B* stated before lead to two content-based grouping strategies: **Content Based Grouping Strategy I.** If list items are coded they can be semantically grouped by applying *Assumption A* forming $C_{ICD-10} = C_{1..i}$ ICD-10 content groups. **Content Based Grouping Strategy II.** By applying *Assumptions A and B*, similar content groups can be formed, with the result that a fraction of the non-coded list items $I_{non\text{-}coded}$ get a code and can therefore be grouped semantically via *Content Based Grouping Strategy I*. At least under *Assumption B*, these non-coded disease list items in $I_{non\text{-}coded}$ can be put together in a way that they are similar in content, via a certain level of *sim* forming $C_{sim} = C_{i+1..n}$ cluster.

This leads to the evaluation of similar content groups (grouping the already coded ones is trivial), of the non-coded disease list items $I_{non\text{-}coded}$ a) by the *Content Based*

Grouping Strategy I (ICD-10 based), so a code could be automatically assigned to a list item in C_{ICD-10} and b) content groups, whereas no code could automatically be assigned to to a list item in C_{sim}. Nevertheless they are automatically grouped due to their string similarity, which reflects the fact of *Content Based Grouping Strategy II* (string similarity based).

We will show that this can be achieved *in one go* using a hierarchical semi-supervised clustering approach where ICD-10 codes are automatically assigned to non-coded list items and at the same time infer the optimal *sim* boundary for string-based list item grouping with a minimal language-depended preprocessing strategy.

3.3. Preprocessing

To normalize the text segments we applied the following Lucene[1] NLP components within a processing chain: a StandardTokenizer for tokenizing the input text; a StandardFilter applying a first Lucene-based text normalization; a LowerCaseFilter, which erases all upper case occurrences; a StopWordFilter[2], removing identified tokens and a SnowballFilter (German2) extracting word stems from tokens. Finally we applied a regular expression for text normalization, removing certain characters from the input token[3] . We refrained from word decompounding even though this is a common German language phenomenon, with certain domain-specific affixes like "-itis" for inflammation or "-ectomie" for surgical removal. Instead, we compensated this language phenomenon with a character n-gram filter, choosing an initial window size of three. This has the side effect that typing errors, an important type of language errors have less impact on token similarity in the VSM (Vector Space Model), explained in the next section.

3.4. Vector Space Model

Applying the VSM model, documents are seen as bag of words and the corresponding terms of a document are mapped to a unique point in a vector space [6, 7]. A document collection D is defined as a set of documents $\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{d}_j$,…,$\mathbf{d}_n$, representing the observations. We did a char 3-gram decomposition of the collection D which forms a vocabulary with $t_1, t_2, t_i, \ldots, t_m$ unique char 3-gram types, which can be interpreted as more fine grained terms in the following descriptions. The document $\mathbf{d}_j$ is spatially interpreted as a point in the m-dimensional vector space, respectively forming an m-dimensional feature vector. Therefore, the VSM is described via an $m \times n$ matrix $\mathbf{X}$. In the classic VSM the term-specific weights within the feature vectors are products of local and global parameters. We use the popular term frequency-inverse document frequency *tf-idf* weighting scheme [8] and the cosine similarity between two documents $\mathbf{d}_i$ and $\mathbf{d}_j$ to obtain the semantic correlation *sim* between two list items I_i and I_j.

3.5. Latent Semantic Analysis

Latent Semantic Analysis (LSA) [9, 10] is a statistical retrieval method exploiting term co-occurrences within a term-document matrix and is generally categorized as a kind of distributional semantics. m terms and n documents form a sparse $m \times n$ matrix $\mathbf{X}$, i.e. the

[1] https://lucene.apache.org/

[2] http://snowball.tartarus.org/algorithms/german/stop.txt

[3] [\d\.\,_\:]+

VSM in Section 3.4. The core of LSA is to apply a singular value decomposition (SVD) on the term-document matrix $\mathbf{X} = \mathbf{TSD}^T$ getting the orthonormal matrices $\mathbf{T}$ and $\mathbf{D}^T$ with the eigenvectors of $\mathbf{XX}^T$ and $\mathbf{X}^T\mathbf{X}$. $\mathbf{T}$ is often called the term matrix and $\mathbf{D}^T$ the document matrix. The roots of the eigenvalues of $\mathbf{XX}^T$ and $\mathbf{X}^T\mathbf{X}$ are embedded in $\mathbf{S}$. The degree of dimensionality reduction can be controlled by eliminating the lowest eigenvalues and their eigenvectors to a new dimension *k*.

3.6. Hierarchical Clustering

In order to provide a content based grouping into *n* sets $C_{1..n}$ we applied a semi-supervised clustering approach. First, we clustered all patient-specific documents $\mathbf{d}_{1..l}$ (50-character long phrases) including the already ICD-10 coded documents with a hierarchical agglomerative cluster method implemented within the *hclust* function provided by the core R programming environment [11]. Agglomerative clustering follows a bottom-up approach in which each document initially is assigned to its own cluster. From this initial configuration, documents are iteratively merged based on a specific distance metric until there is just a single cluster. UPGMA for instance, which was used in this work, computes the distances between two cluster A and B based on the pairwise average distances between their assigned documents $\mathbf{d}$.

$$UPGMA = \frac{1}{|A||B|} \sum_{\mathbf{d}_a \in A} \sum_{\mathbf{d}_b \in B} f_{sim}(\mathbf{d}_a, \mathbf{d}_b) \tag{1}$$

Motivated by the hypotheses that string similarity of diagnosis phrases correlates with ICD-10 assignments (Section 3.2 *Assumptions A and B*) we applied different cut heights to the resulting tree diagram. In this particular step, we inferred the best cut-off in order to reproduce the already coded ICD-10 clustering scheme (I_{coded}) judged by the F-measure [12]. Based on the resulting groups, unlabeled documents ($I_{non\text{-}coded}$) were coded if and only if they appeared in a cluster together with at least one ICD coded document. If two documents with different ICD-10 codes were clustered in the same group, we assigned the label of the document with the smallest cosine distance.

4. Results and Discussion

The initial clustering based on string similarity to reproduce the ICD-10 clusters scored an F-measure of 0.71 on our test data set. This result suggests that string similarity generally can be assumed to correlate with ICD-10 encodings and thereby supports our hypotheses (*Assumption B*). However, while we observe a high precision of 0.82 we get a relatively low recall of 0.62 (Figure 1a). Thus, our algorithm tends to separate documents into different cluster although these documents belong to the same ICD-10 group. For instance, the two diagnoses "NTX"[1] and "Niereninsuffizienz dialysepflichtig" (renal failure requiring haemodialysis) are assigned to the same ICD-10 code (N18) by the physician, but are clustered into different groups based on the hierarchical clustering algorithm (*false negative*). However, the per-patient post ICD description assignment introduced in this paper naturally corrects for this scenarios due to the following feature:

[1] NTX is an abbreviation for "Nierentransplantation" (kidney transplant)

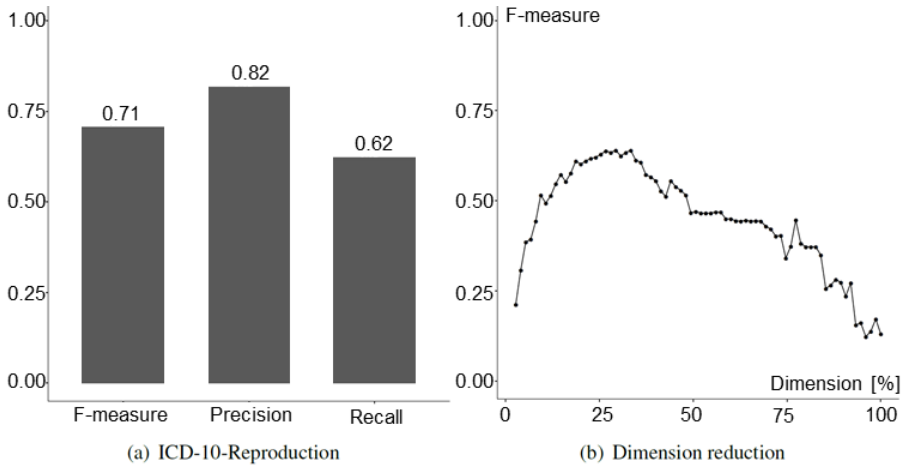


Figure 1. (a) The power of hierarchical clustering using string similarity to reproduce the ICD-10 groupings I_{coded} is demonstrated. The string similarity approximately correlates with the ICD-10 assignments with a moderate F-measure and a good precision value. At the same time, it highlights the need of further information about string patterns within ICD-10 cluster for per-patient post ICD description grouping. (b) We used LSA to reduce the feature space. For each dimension-reduction step, we cut the tree with different heights and plotted the maximum F-measure achieved for the particular subspace.

As long as both documents are clustered together with a same ICD-10 coded document, both will simultaneously get the same ICD-10 code even if their string dissimilarity is high.

Using LSA we were able to reduce the feature space up to 75 percent with just a small loss of accuracy (Figure 1b). The linear transformation into a new semantic space via LSA increased not yet the accuracy, but could be an advantage exploiting a large scale approach with thousands of patients in the future. To evaluate our method we made two assumptions. First, we assumed that the physicians correctly assigned ICD-10 codes to the documents and second, we ignored ICD-10 codes that were not used already for a specific patient. Documents that were not coded prior to the analysis and get the ICD-10 code are defined as *true positives*, documents that were not coded and got the wrong ICD-10 code are *false positives* and finally, documents with missing ICD-10 codes but should have get one are *false negatives*. Based on this evaluation framework we calculated the F-measure to measure the performance of our algorithm (Table 1). In addition, we repeated the evaluation and modified our second assumption. In this case, all ICD-10 codes are assumed available, resulting overall in a much higher *false negative rate* (Table 1, numbers in brackets). However, it is likely that ICD-10 codes not yet assigned to a specific patient might be available in documents related to other patients. Therefore, considering a large amount of patients we expect a substantially reduced *false negative rate* in this scenario. Given that and the results shown in Table 1 we are expecting an F-measure of > 0.9 in post-assigning ICD-10 codes in a large scale analysis.

Table 1. Post ICD-10 coding* and string clustering results**

Patient	Precision	Recall	F_1*	Precision	Recall	F_1**
P_1	0.93	0.96 (0.74)	0.94 (0.83)	0.78	1.00	0.88
P_2	0.90	0.82 (0.61)	0.86 (0.73)	0.91	0.78	0.84
P_3	0.73	0.98 (0.69)	0.83 (0.71)	0.84	0.81	0.82
P_4	0.91	0.97 (0.87)	0.94 (0.89)	1.00	1.00	1.00
P_5	0.80	0.90 (0.63)	0.85 (0.70)	0.59	0.93	0.72

Table 2. Number of the identified topics out of the initial disease list items.

Patient	List items	Unique list items	Topics	Compression rate
P_1	302	184	60	0.80
P_2	250	174	70	0.72
P_3	861	441	95	0.89
P_4	531	295	77	0.85
P_5	378	262	90	0.76

A notable amount of documents might be not assignable to a certain ICD-10 code, but the algorithm proposed in this work categorizes those remaining non-coded documents based on string similarity (*Assumption B*) therefore we achieve a good topic clustering rate for *all initial diagnosis list items* (Table 2). The method suggests good performance in four patients, properly detecting the disjoint content classes of the diagnoses. Patient 5, in contrast, scored a low precision value of 0.59 due to a low cosine cut-off value learned from the already ICD-10 coded documents I_{coded} causing a high *false positive rate*.

5. Conclusion

In this paper, we showed that diagnosis or problem lists in EHRs can be condensed by a clustering approach. This suggests the feasibility of a first patient QuickView implementation with the goal to enhance the readability of long clinical diagnosis or problem listings. However, the results also point to a set of scenarios, which remain challenging to address. a) Physicians often put multiple diagnoses within a single phrase, which makes it hard to classify for clustering algorithms where a document cannot be assigned to two different clusters at the same time. b) In some cases the codes were plainly incorrect. As we are learning from these coding schemes this can become a real issue in categorizations of a small amount of available list items. In general, we expect that the majority of physicians are choosing the right codes, and we intend to make use of this realistic assumption in future work. c) Abbreviations are still difficult to handle although the problem minimizes when a physician already assigned an ICD-10 code for the abbreviation as well as for the corresponding full text. We hypothesize that by exploiting a larger patient-specific database we can address these problems, by raising the documentation quality for the individual patient out of many patients, which we want to investigate in the future.

Acknowledgements

This work is part of the IICCAB project (Innovative Use of Information for Clinical Care and Biomarker Research) within the K1 COMET Competence Center CBmed (http://cbmed.at), funded by the Federal Ministry of Transport, Innovation and Technology (BMVIT); the Federal Ministry of Science, Research and Economy (BMWFW); Land Steiermark (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the FFG.

References

[1] B. Koopman, G. Zuccon, A. Nguyen, A. Bergheim, and N. Grayson. Automatic ICD-10 classification of cancers from free-text death certificates. *Int J Med Inform*, 84(11):956–965, Nov 2015.
[2] B. Koopman, S. Karimi, A. Nguyen, R. McGuire, D. Muscatello, M. Kemp, D. Truran, M. Zhang, and S. Thackway. Automatic classification of diseases from free-text death certificates for real-time surveillance. *BMC Med Inform Decis Mak*, 15:53, Jul 2015.
[3] W. Ning, M. Yu, and R. Zhang. A hierarchical method to automatically encode Chinese diagnoses through semantic similarity estimation. *BMC Med Inform Decis Mak*, 16:30, Mar 2016.
[4] Y. Chen, H. Lu and L. Li, Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity. *PloS one*, 12(3):e0173410, 2017.
[5] S. Boytcheva. Automatic matching of ICD-10 codes to diagnoses in discharge letters. In *Proceedings of the Workshop on Biomedical Natural Language Processing*, volume 9, pages 11–18, 2011.
[6] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):620, 1975.
[7] E. Boerjesson and C. Hofsten. A vector model for perceived object rotation and translation in space. *Psychological Research*, 38(2):209–230, 1975.
[8] R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*. Addison- Wesley Reading, MA, 1999.
[9] T. K. Landauer and S. T. Dumais. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240, 1997.
[10] T. K. Landauer, P. W. Foltz, and D. Laham. An introduction to latent semantic analysis. *Discourse Processes*, 25:259–284, 1998.
[11] R Core Team. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
[12] C. D. Manning, P. Raghavan, H. Schütze, et al. *Introduction to information retrieval*, Volume 1. Cambridge University Press Cambridge, 2008.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-108

Modular Architecture for Integrated Model-Based Decision Support

Jan GAEBEL[a,1], Erik SCHREIBER[a], Alexander OESER[a] and Steffen OELTZE-JAFRA[a]
[a] *Innovation Center Computer Assisted Surgery (ICCAS), Faculty of Medicine, University Leipzig, Leipzig, Germany*

Abstract. Model-based decision support systems promise to be a valuable addition to oncological treatments and the implementation of personalized therapies. For the integration and sharing of decision models, the involved systems must be able to communicate with each other. In this paper, we propose a modularized architecture of dedicated systems for the integration of probabilistic decision models into existing hospital environments. These systems interconnect via web services and provide model sharing and processing capabilities for clinical information systems. Along the lines of IHE integration profiles from other disciplines and the meaningful reuse of routinely recorded patient data, our approach aims for the seamless integration of decision models into hospital infrastructure and the physicians' daily work.

Keywords. Clinical Decision Support Systems, Decision Modeling, Computerized Models, Computer Architecture, Model Sharing.

1. Introduction

Clinical decision support systems (CDSS) have great potential to bring added value to the daily clinical practice. They are supposed to support physicians in diagnosing or finding a better, more individualized therapy for complex diseases, e.g. head and neck cancer [1,2]. Developing patient models as a foundation for CDSS is a suitable tool to achieve personalized medicine [3]. Patient models can be utilized to integrate patient-specific data from hospital information systems (HIS). Thereby, they can help to calculate and assess the respective situation, e.g. possible therapy options given specific patient characteristics. The application of these models, however, would require a substantial infrastructure of dedicated systems.

In general, a typical HIS does not support the sharing or processing of any model-based functionality on the basis of patient-specific data. Although additional storage for general or patient-specific models can be realized, the processing capability needs to be added with respective software components. To avoid additional efforts for the physicians in using a CDSS, the interoperability of the respective subsystems must be considered. There are also several different settings where decision models in the context of clinical decision support might be used. For an oncological example, functionalities of a CDSS could be used on a physician's workstation during his or her clinic hours or by an interdisciplinary team of physicians in a tumor board meeting. Therefore, isolated installations on only a few clinical workstations should be avoided [4]. To be able to

[1] Corresponding Author: Jan Gaebel, ICCAS, Faculty of Medicine, University Leipzig, Semmelweisstraße 14, 04103 Leipzig, Germany, E-Mail: jan.gaebel@medizin.uni-leipzig.de.

create a sophisticated, institution-wide solution (or even beyond the boundaries of one institution), there is the necessity to exchange the models and their respective processing engines. They must be available at every other point of use. Reuse and exchange of data in clinical environments is a continuously investigated subject [5]. The model exchange, as the foundation for a well-integrated model-based decision support, must be based on well-established clinical standards. This will facilitate interoperability with other systems, a seamless integration into existing clinical IT landscapes and easy adoption. In this paper, we propose an architecture of dedicated subsystems for the efficient integration and standardized exchangeability of decision models in a HIS environment.

2. Methods

A HIS, typically running on a modern computer network, provides the basic infrastructure on top of which a CDSS is built [6]. Administrative patient data and clinical data are stored in various databases within the HIS network. They are stored in a safe way, fulfilling all legal requirements and demands on retrieval of all persons involved. Typically, functionalities of a HIS do not go beyond these requirements. Juridical and license regulations aside, these data can however be accessed in the space of the network. To avoid any redundant data storages, the routinely used databases should be used by any additional systems for any added value as well. The network infrastructure generally allows for the use of modern communication techniques. The connection of several interlinked systems via web technologies, e.g. HL7 FHIR provides for semantically structured storage and RESTful services to transfer the data [7].
For a specific analysis of exemplary integration scenarios, we chose to analyze the decision model for laryngeal cancer, proposed by Cypko et al. and Lemke et al. in [8,9] and deduced a list of more generalized requirements. For this type of decision support, Figure 1 illustrates the requirements that must be addressed and fulfilled by the model-based CDSS:

Figure 1. Three categories of requirements for model integration

- Interaction with models through various user interfaces
- Easy storage, access and exchangeability of models
- Maintenance and revision control of models
- Instantiation of models with patient-specific data
- Reuse of routinely recorded data with minimum additional efforts for maximum efficiency

Different users with different levels of experience and needs will be using decision support systems. Systems, specifically for support in oncology, need to provide various viewpoints. Along the clinical pathway of laryngeal cancer treatment, there are several stages where such a system would come into use. For each purpose, different user interfaces and visualizations are needed. However, all frontends need to be built on the same model base and need to use the same engines and infrastructure. This guarantees consistent support and results presentation as well as a high user acceptance. Possible scenarios supported by a CDSS in head and neck oncology, which will be enabled by the proposed architecture, are:

- the attending physician analyzes a complex patient case in preparation of the interdisciplinary tumor board meeting
- multiple physicians discuss the current patient at the interdisciplinary tumor board meeting and develop a treatment strategy
- when discussing the planned treatment strategy, both patient and physician require a suitable presentation of the patient's situation
- during treatment, for example the surgeon is assisted in real-time on how to proceed with the intervention given an unexpected event

In general, the following dedicated systems would be required to establish an integrated system for processing and exchanging functionalities of model-based decision support. The main goals of such a modular system are a) to get the best available technologies for the functionalities of the individual systems, b) exchange their individual functionalities via modern information and communication technologies, and c) to be able to integrate the decision support functionality into existing hospital information systems with interchangeable interfaces to clinical databases or user frontends. The system would consist of:

(1) a processing unit for the model-based calculations
(2) storage systems for different kinds of models
(3) access to and preparation of routinely recorded patient data
(4) user interfaces

The processing unit (1) utilizes model instantiation and mathematical inferencing techniques. It utilizes provided general models from the storage system (2). They may formally represent specific disease or general patient characteristics and their relations. These models can be instantiated into the patient-specific models with data from clinical databases (3). In this way, assessment and inference of desired information, e.g. possible treatment options for a given patient, are possible. These results must be presented to the decision maker (4). The central processing unit must be connected to all the other subsystems. It listens and responds to their requests and event calls.

The model storage consists of two parts. Firstly, general decision models need to be stored and hold available. To support maintenance, distribution and content-related adjustments (e.g. adjustments according to updated clinical guidelines), models need to be located in a model distribution system, containing multiple model repositories depending on the specific use cases, separated from the processing engine [10]. This

allows the exchange or adding of individual repositories. This might include the expansion of the distribution system with more repositories spread amongst different institutions. For simulating and testing purposes, instantiated models can be used with specific data elements set to desired values which need to be stored as well. Revision control is necessary to allow for multiple versions of simulated patient situations. These patient specific models must be accessible for comparison and planning.

Depending on the respective HIS, a connecting component is needed to manage the data access. In general decision models, the necessary information entities are formally defined. These entities must be instantiated with patient specific information to be able to calculate and infer the desired information, e.g. therapy options. Therefore, different kinds of preprocessing could be necessary. On the one hand, a structural preprocessing must be done. Patient information is mostly stored in unstructured, narrative texts. The information can be extracted using natural language processing. On the other hand, a semantic preprocessing is important to get reliable and valid patient specific information into the processing chain. An example might be an evaluation of temporal characteristics of the data, as proposed by Gaebel et al. in [11].

For the different application scenarios listed above, the system architecture must provide interfaces for respective user frontends. This is realized by providing request/retrieve functionalities for models via data communication on well-established standards. This must ensure the communication of patient specific models with all associated patient data and further meta-information between subsystems.

3. Results

We propose an architecture concerning the mentioned use case of a decision model for laryngeal cancer therapy [8]. It utilizes Bayesian networks to formalize the relations of information of a patient with laryngeal cancer. Nodes inside the network can be instantiated with patient-specific data from clinical observations. Based on evidence-based conditional probabilities inside the network, the Bayesian inference algorithm allows the approximation of unknown information. This is used to infer possible patient-specific options for a personalized therapy. In the following, we describe our approach to integrate this model-based decision support into existing information system environments by addressing the identified requirements.

We built the model processing engine as the central dedicated system for our architecture. All core components are implemented in JAVA. Other components or sub systems are integrated via web services (REST, SOAP). We either implemented the additional systems ourselves or resorted to existing solutions. Figure 2 illustrates the system infrastructure with the different sub systems: 1) Central Processing engine, 2) Model Distribution System, 3) Data Access & Preprocessing engine and 4) User Interfaces & Frontends.

We built the central engine as a server application that provides its functionalities as web services for every other system. We implemented it using Java Spring Boot for easy configuration and functional security. The processing of the patient models is implemented using the SMILE engine[1], a framework for developing and processing Bayesian networks. Patient models are requested via web services from the model distribution system. It is built on the basis of the cross-enterprise model sharing

[1] BayesFusion, LLC, https://www.bayesfusion.com/smile-engine

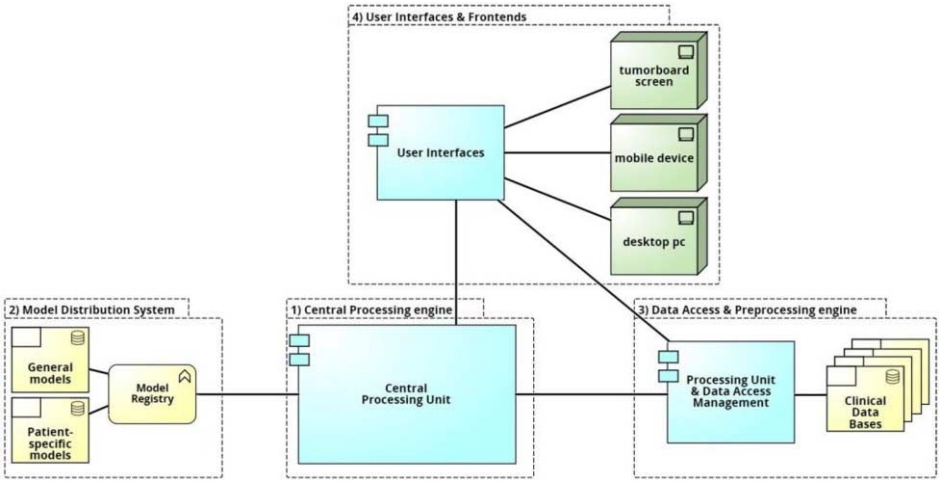


Figure 2. Modular architecture for the integration of probabilistic decision models

architecture (XMS) [10], which provides modular, standardized model distribution functionalities and is currently a work item of the IHE Surgery Domain [12]. Retrieved patient models are instantiated with patient specific data from the clinical database. The central engine comprises all necessary functions to handle the patient models. For decision support, all unknown states of the patient (e.g. suitable therapy options) are inferred. This inference is based on the intrinsic structure of the respective models and their conditional probabilities, as stated by Cypko et al. in [8].

For further analysis, different methods of investigating a Bayesian patient model are provided. For instance, information entities with a strong influence on therapy decisions can be highlighted and a sensitivity analysis of predicted values can be conducted. The results of these operations are also supplied as web services. Bayesian network models allow for simulating different scenarios, e.g. distinct treatment options with their respective outcome. For this purpose, certain information entities are set manually, e.g. surgery as the desired treatment option. Subsequently, the respective information about the outcome is inferred. Hence, several different scenarios can be simulated and evaluated. These different states of a patient-specific network must also be stored and retrieved, to avoid the need of reconfiguration every time the simulated outcomes will be assessed. The SMILE engine allows for these configurations. Instantiated or simulated models are passed to the model distribution system in a proprietary xml-type format.

The repositories for general and patient-specific models are implemented using a H2 database server[1]. Therapy decision models are stored in an xml-type file format. It is a proprietary format necessary for the SMILE engine to process the models. Different models are identified and requested by a unique identifier. When instantiated and stored as patient specific models, they are then identified and requested by unique model identifier, unique patient id and version number. For testing purposes, we are using anonymized patient data provided by our clinical partners which is stored with an arbitrary identifier. A model registry is implemented to receive the requests and point to the correct repository and/or database.

[1] http://www.h2database.com/

In our proposed system, we installed an ArdenSuite-Server[1] to process incoming data requests. The server provides functionalities of Arden Syntax, a programming language specifically designed for medical application and decision support, as web services. We implemented Arden Syntax medical logic modules (MLM) to handle the preprocessing. Incoming requests for specific patient data are forwarded to a MySQL database, which existed from other projects for testing purposes and where anonymized patient data are stored. These data are identified by an arbitrary patient id. We make use of Arden Syntax's curly brace expressions [13] to define specific queries for a patient's information, e.g. findings from various diagnostic procedures. The response data is received by the respective MLM, where it can be semantically preprocessed. The validity of the data can be evaluated or necessary scores can be calculated by the MLM. The requested information with additional values, e.g. scores or quality measures, are then sent back to the central engine in a formatted JSON string.

The integration of model-based decision support requires novel user interaction. Different devices for different use cases run different frontends, e.g. applications on a physician's workstation [14], shared decision-making in an outpatient setting [15] or large dashboard screens in a meeting room. But with the integration of the central processing unit via web services, it can be ensured that all results and decision support functionalities can be applied whenever and wherever they are needed. Reusability of the data also ensures an efficient and safe application of the decision support system.

The described system is in use at the Innovation Center Computer Assisted Surgery, Medical Faculty, University of Leipzig, Germany. It provides the foundation for a research project in model-based clinical decision support. The illustrated subsystems are used to test and further develop therapy decision support for an exemplary clinical use case of laryngeal cancer. In close collaboration with head and neck surgeons, we are processing several anonymized test cases of patients with laryngeal cancer to improve the results of the individual subsystems [8]. Application systems with novel interaction facilities are being developed to provide access to and better understanding of the decision models [14].

4. Discussion

In this paper, we propose a modular architecture for the integration of model-based decision support into existing clinical environments. We focused on the requirement of sharing the underlying decision models and the integration of patient-specific data. A specific use case from current scientific research was used to derive general requirements. A reference implementation was described.

The individual components within the introduced architecture can be adjusted to the existing IT infrastructures or different functional needs of the respective department. Layout and design of the front end components should be individually adapted to the particular user. But there are also alternatives to the previously mentioned back end solutions.

The repository storage of the decision models is currently handled by a relational database storing the model in a text-based xml-format. Since the contained data represents a complex structure of interlinked patient information, the files are hard to interpret in a human-readable form. An alternative way to store the model data is the use

[1] http://www.medexter.com/products/ardensuite

of a relational schema adjusted to the model structure. Individual model components, nodes of the Bayesian network in our use case, could be stored and queried as individual database items connected by using a sophisticated database schema. However, relational databases are not particularly suited for this specific task, since the complex structure of the probabilistic graphical model needs to be converted into a relational interpretation. This approach results in a massive growth in schema size and the amount of complex queries to get values for subsequent systems. However, non-relational databases like OrientDB or Neo4j are specifically designed to store graph-based data. In combination with dedicated database management, the overall infrastructure can benefit from features like user roles, parallel access, automated versioning as well as consistency and integrity checks.

The current version of the central processing unit is equipped with the SMILE engine because of its suitable feature set in the calculation of Bayesian networks for probabilistic reasoning. However, the infrastructure is adaptable to other processing engines like e.g. Bayes Server or any other application that offers a similar feature set as well as an appropriate API (Application Programming Interface) for data handling based on the specific characteristics of the respective model type.

A crucial problem in utilizing patient information is their extraction from narrative texts. Developers of decision support systems must either use natural language processing methods that usually lack high-precision or rely on additional input methods. In this case, physicians need to enter the structured patient data manually into the new system. Vendors are trying to integrate more structure into their information systems, e.g. storing patient data as FHIR resources [7], but it will take a long time until the majority of HIS offer such technologies.

For the maintenance of the general disease models, additional methodologies are needed. Depending on the creation method (e.g. manual model creation vs. machine learning) different strategies might be appropriate. But medical knowledge is changing. Hence, the models need to be kept up-to-date [16]. There are several approaches to easily let the domain-experts adapt and maintain the models themselves. These systems would have to interact with the model distribution systems as well as the central processing unit. This functionality has been considered, but not yet been realized in our infrastructure.

References

[1] Lucas PJF, Boot H, Taal BG. Computer-based decision-support in the management of primary gastric non-Hodgkin lymphoma. Meth Inform Med 1998;37:206–19.
[2] Oniśko A, Druzdzel MJ, Wasyluk H. Extension of the Hepar II Model to Multiple-Disorder Diagnosis. In: Intelligent Information Systems. Advances in Soft Computing, vol 4. Physica, Heidelberg. 2002.
[3] Chan IS, Ginsburg GS. Personalized Medicine: Progress and Promise. Annual Review of Genomics and Human Genetics. 2011;12(1):21744.
[4] Schuh C, de Bruin JS, Seeling W. Clinical decision support systems at the Vienna General Hospital using Arden Syntax: Design, implementation, and integration. Artif Intell Med. 2015 Dec 1;
[5] Murphy EV. Clinical decision support: effectiveness in improving quality processes and clinical outcomes and factors that may influence success. Yale J Biol Med. 2014 Jun;87(2):187–97.
[6] Winter A, Haux R, Ammenwerth E, Brigl B, Hellrung N, Jahn F. Health Information Systems: Architectures and Strategies. Springer Science & Business Media. 2011.
[7] FHIR v3.0.1 [Internet]. [cited 2018 Jan 10]. https://www.hl7.org/fhir/
[8] Cypko MA, Stoehr M, Kozniewski M, Druzdzel MJ, Dietz A, Berliner L, et al. Validation workflow for a clinical Bayesian network model in multidisciplinary decision making in head and neck oncology treatment. International Journal of Computer Assisted Radiology and Surgery. 2017;112.

[9] Lemke HU, Cypko MA, Berliner L. Information integration for patient specific modelling using MEBNs: example of laryngeal carcinoma. In: Proceedings of the Int Joint Conf CARS. 2012. p. 2468.
[10] Schreiber E, Franke S, Meier J, Neumuth T. Towards a surgical process model distribution infrastructure based on Cross-Enterprise Model Sharing (XMS). In: Proceeding of the 29th Conference for Comput Assist Radiol Surg. Barcelona, Spain. 2015.
[11] Gaebel J, Cypko MA, Oeltze-Jafra S. Considering Information Up-to-Dateness to Increase the Accuracy of Therapy Decision Support Systems. Stud Health Technol Inform. 2017;243:21721.
[12] IHE Surgery Domain Page: Brief Proposals 2017/2018. http://wiki.ihe.net/index.php/Surgery_Proposals_2017-2018#Brief_Proposals. [cited 2018 Jan 31]
[13] Kraus S, Enders M, Prokosch H-U, Castellanos I, Lenz R, Sedlmayr M. Accessing complex patient data from Arden Syntax Medical Logic Modules. Artificial Intelligence in Medicine. 2015 Sep;S0933-3657(15):120–7.
[14] Cypko MA, Wojdziak J, Stoehr M, Kirchner B, Preim B, Dietz A, Lemke HU, Oeltze-Jafra S. Visual Verification of Cancer Staging for Therapy Decision Support. Computer Graphics Forum. 2017 Jun 1;36(3):109–20.
[15] Müller J, Zebralla V, Wiegand S, Oeltze-Jafra S. Interactive Visualization of Functional Aspects in Head and Neck Cancer Aftercare. In IEEE VIS, Phoenix, Arizona; 2017.
[16] Cypko MA, Oeltze-Jafra S, Stoehr M, Dietz A, Berliner L, Lemke HU. Quality Management of Therapy Decision Models based on Bayesian Networks. In: Proceeding of the 31st Conference for Comput Assist Radiol Surg. 2017.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-116

Cleansing and Imputation of Body Mass Index Data and Its Impact on a Machine Learning Based Prediction Model

Stefanie JAUK[a,1], Diether KRAMER[b] and Werner LEODOLTER[b]
[a] *CBmed, Graz, Austria*
[b] *Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria*

Abstract. *Background:* A challenge of using electronic health records for secondary analyses is data quality. Body mass index (BMI) is an important predictor for various diseases but often not documented properly. *Objectives:* The aim of our study is to perform data cleansing on BMI values and to find the best method for an imputation of missing values in order to increase data quality. Further, we want to assess the effect of changes in data quality on the performance of a prediction model based on machine learning. *Methods:* After data cleansing on BMI data, we compared machine learning methods and statistical methods in their accuracy of imputed values using the root mean square error. In a second step, we used three variations of BMI data as a training set for a model predicting the occurrence of delirium. *Results:* Neural network and linear regression models performed best for imputation. There were no changes in model performance for different BMI input data. *Conclusion:* Although data quality issues may lead to biases, it does not always affect performance of secondary analyses.

Keywords. Electronic health records, body mass index, machine learning, data imputation, data cleansing, predictive modelling.

1. Introduction

Within the last years, several prediction models based on machine learning (ML) methods have been published, many of them based on data from electronic health records (EHRs) [1,2]. The advantage of using EHRs as an input for prediction modelling is the big amount of collected longitudinal data, which can be used as training and validation sets [3].

Although the use of EHRs may be beneficial for prediction models, challenges for the secondary use of data need to be considered. Criticism of using EHRs in secondary analyses is often centred on data biases and quality issues. Not only incomplete but also incorrect data may affect the outcome of prediction models. Cruz and Wishart [4] stated that if data used for machine learning is of poor quality the results will be as well. The term "garbage in, garbage out" is commonly used to describe this scenario. Hence, increasing the quality of the data used as input for prediction models may improve model performance and lead to more accurate prediction.

One example of treating missing data in EHR is the work of Jerez et al. [5]. The authors compared different imputation methods in a data set consisting of 3,679 records for modelling early breast cancer relapse. Six artificial neural networks were trained

[1] Corresponding Author: Stefanie Jauk, CBmed GmbH – Center for Biomarker Research in Medicine, Stiftingtalstrasse 5, 8010 Graz, Austria; E-mail: stefanie.jauk@cbmed.at.

using different imputation methods and prognostic accuracies of all models were compared. However, as their aim was to improve the accuracy of the prediction model, they did not evaluate the quality of imputed values but only final accuracy.

While some data in EHRs are missing at random, other records include systematically missing values which may lead to biases. An example for the latter case is body mass index (BMI). BMI is more likely to be measured in patients suffering obesity or anorexia, and records are more likely to be missing in patients within the normal range of BMI [6]. Recording of BMI is also correlated with disease status, e.g. BMI data is more frequent for patients with diabetes mellitus than for those without [7]. Apart from recording bias, reporting biases may exist in BMI data if self-reported, e.g. more missing values for women who categorized themselves as overweight or underweight [8].

In healthcare, BMI data is an important parameter for risk assessment, such as for fractures [9], heart failure [10] or preeclampsia [11]. Besides, BMI data has been used as a modelling feature in several risk prediction models developed with ML. An example is the work of Kramer et al. [12] who developed a model predicting the occurrence of delirium in hospitalized patients. Further investigations of the developed model showed that BMI was one of the variables with highest importance. Furthermore, malnutrition has been identified as a risk factor for delirium [13] and we assume that a low BMI can be used as an indicator.

Due to its importance for risk prediction and keeping in mind the biases of BMI recording, BMI cleansing and imputation methods are essential though challenging. A common problem of clinical data cleansing is the observation of extreme values. Cleansing of clinical BMI data can be difficult, as distributions for weight and height may differ from healthy samples. Freedman et al. [14] illustrated that plausibility intervals for weight and height like the ones set by WHO often result in numerous outliers which are set to missing, even though the values are correct.

Facing the problem of missing data in EHRs, Kontopantelis et al. [6] developed a multiple imputation algorithm for longitudinal body mass index data. Their aim was to produce imputation values for variables with very low individual variability. However, their newly established algorithm did not always perform better than the reference algorithm and appeared to have very long computation times. Another limitation of their work is its applicability on EHRs of hospital information systems (HIS). Many patients do not have any BMI documented in their EHR. In these cases, the algorithm cannot be used as it ignores patients with less than two BMI values.

In order to improve the quality of the secondary use of EHRs, we wanted to investigate and improve the shortcomings of BMI data on routine data of a HIS. The data for analyses belongs to Steiermärkische Krankenanstaltengesellschaft m.b.H (KAGes), the regional health care provider in Styria (Austria). The HIS of KAGes hosts longitudinal health records of around 90 % of all Styrian inhabitants, including hospital stays and outpatient visits over 15 years [12].

Hence, the aim of this study is to (1) establish an automatic method for data cleansing and imputation of missing values for BMI data of a hospital information system, and (2) assess the impact of cleansing and imputation on the performance of an already established prediction model.

In a first step, we evaluate common sources of incorrect values for weight and height due to data entry errors and predefine intervals for plausible values. After cleansing of existing BMI values, missing values will be addressed. In contrast to Kontopantelis et al. [6] we believe that even though height will be quite stable over time, weight

recordings may undergo changes, depending on a patient's health status. In addition, there may be good indicators for extreme values, e.g. correlations between weight and diseases like anorexia and obesity. Imputation approaches may reflect those correlations and result in accurate imputed values. In contrast to the work of Jerez et al. [5] our aim is to obtain imputations closest to real data, not imputation methods that perform best in prediction modelling. If quality of clinical data can be influenced via a data cleansing and imputation step, we expect a change in the accuracy of risk prediction. Therefore, we will assess whether such a step improves the performance of a prediction model.

2. Methods

2.1. Information Extraction

All information was extracted from the KAGes hospital information system *openMEDOCS,* based on IS-H/i.s.h.med information systems and implemented on SAP platforms. In openMEDOCS, height and weight values are recorded via nursing assessment. Patients self-report their current values and nursery personnel enter the numbers in a free-text entry field. BMI values are automatically calculated once height and weight values are available. As long as height and weight are self-reported and not measured, we assume reporting biases in the data.

Between 2011 and 2017 753,701 patients have had at least one admission in a KAGes hospital. Out of those, 39,015 (5.2 %) patients had only height data and 5,300 (0.7 %) only weight. For 346,010 (45.9 %) patients both values were available.

Before extracting lab data and diagnoses for all patients, we defined inclusion criteria: Patients must have at least one admission between 2011 and 2017, height and weight recordings must be available, and age must be over 18 years. The sample resulted in 328,283 patients, with a total of 709,003 admissions. The median BMI for all patients was 25.88 kg/m², with a lower quartile of 22.89 kg/m² and an upper quartile of 29.30 kg/m².

2.2. Data Cleansing

As longitudinal records were available for some patients, we cleansed the data for each admission and then extracted the latest plausible information for each patient.

First, we defined plausible ranges for height and weight data in hospitalized patients. Considering the research of Freedman et al. [14] we chose ample intervals to include correct outliers. For weight and height values we allowed the intervals [20 kg, 250 kg] and [120 cm, 250 cm], respectively.

Second, we analysed the extracted data and discovered several systematic errors that most probably occurred during data entry. We adjusted the implausible values due to these errors, before applying the plausibility intervals for cleansing:

- A small number of EHRs had height values within plausible ranges, but values for weight over 400. We assumed this to be due to a missing decimal point and multiplied weight values by 0.1.
- Some EHRs showed weight within plausible ranges, but height values between 12 and 20. The most probable reason for such values is a missing last digit and multiplied height by 10.

- A larger group of EHRs showed height values below 130 and weight above 130 at one admission. After a visual examination we considered those records to be inverted and simply swapped them.
- In some EHRs we found plausible weight values, but values for height ranged between 50 and 99. Again, we assumed this to result because of a missing digit and corrected this group of records by adding 100 centimetres.

KAGes has already addressed some of those in a data entry control function. For implausible BMI values an alert is triggered during the documentation process. Nevertheless, errors are still possible and previous data needs to be corrected in order to use it for the training of prediction models.

2.3. Imputation of Missing Height and Weight Values

One aim of our study was to compare different imputation methods and to choose the best for further analyses on prediction modelling. We extracted a random sample of 40,000 hospital admissions corresponding to 30,104 patients for whom we compared several statistical methods and machine learning methods. Among possible statistical imputation methods we chose the median imputation, a linear regression model and a linear regression model based on multiple imputation. Machine learning uses the k-nearest neighbor method (KNN), a random forest (RF), a neural network (NN) and a support vector machine (SVM).

All imputation analyses were carried out in R. For the linear regression model with multiple imputation we used the mice package [15]. KNN was modelled with the VIM package and the SVM with the e1071 package. For RF and NN we used the caret package [16] with a 10-fold cross validation as recommend by Kuhn and Johnson [17]. The NN was modelled inside the caret package with nnet, using a feed-forward neural network with a single hidden layer.

We used 75 % of the data for training and tested the results on the remaining 25 % of the data. The training set consisted of 114 features for prediction, including gender, diagnoses and lab data. The features were selected during pre-processing using bivariate analytics. After modelling, we compared the predicted values for the test data set with the real values obtained from the EHR. In order to determine the precision of predictions, we computed the root mean square error (RMSE) for the predictors of BMI between all methods. The RMSE represents the square root of the average of the differences between the real values and the estimated ones.

2.4. Effects of Imputation on a Prediction Model

In order to examine the effect of cleansing and imputation of BMI values on a prediction model we used an already defined model predicting an occurrence of delirium in hospitalized patients. The sample for modelling resulted in 29,568 patients with or without an occurrence of delirium. Out of those, 13,325 patients (45.1 %) had no BMI values. A random forest with down-sampling was used as a classification method and modelled with the caret package in R [16]. The feature set was based on 321 features, including demographic data, ICD-10 diagnoses codes, lab data and BMI. More details of modelling and feature selection can be found in [12].

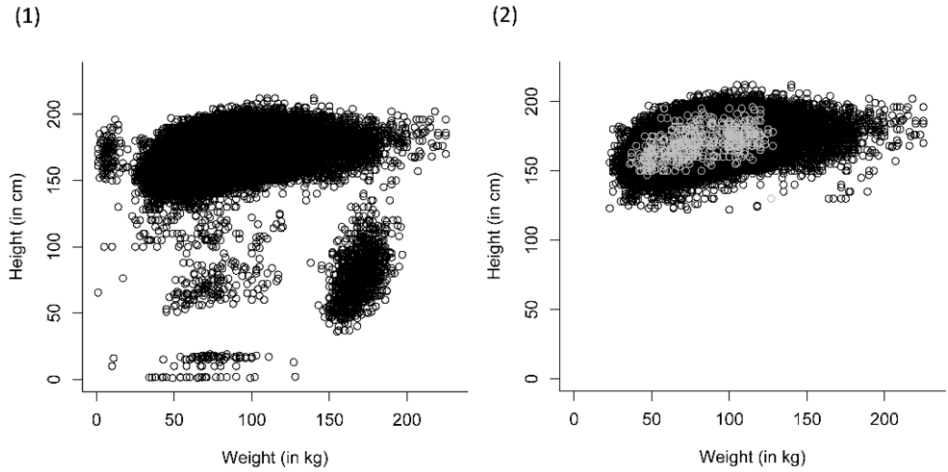


Figure 1. Height and weight records for 709,003 EHRs of adults before (1) and after (2) data cleansing. In (2), values that were affected by data cleansing are highlighted in grey.

In a first scenario, the prediction model was trained and evaluated without including BMI data. For a second simulation, we used the BMI records obtained from the HIS (including wrong or missing values) as a feature. In a third modelling process, the cleansed and imputed BMI values were used as BMI feature. Performance of the prediction model was evaluated using accuracy, sensitivity and specificity.

3. Results

3.1. Data Cleansing of Height and Weight Values

Results of performing the cleansing method on the total sample are shown in Figure 1, comparing original height and weight records and cleansed data.

The group of EHRs with weight values above 130 kg, and height below 130 cm needs to be highlighted. These values build a cluster which is due to the exchange of height and weight entry. The cleansing affected 2,077 records in total. Apparently incorrect values due to data entry were successfully cleansed, as shown in Figure 1.2 with grey data points.

3.2. Comparison of Imputation Methods

To assess the precision of the different imputation methods, we compared the predicted values of all methods with the values obtained from the EHR data for the test data set (Table 1). As expected, computation times were much higher for machine learning methods than for statistical methods. The lowest RMSE was achieved by the NN method with 4.34. However, computation time for NN was the highest with more than ten hours. Results for the linear regression model with simple imputation and the one with multiple imputation were comparable to the NN results, but with computation times below 30 seconds We compared scatter plots of all methods (excluding median imputation) for original and fitted values (Figure 2). For further analyses, we used the linear model with simple imputation as model for imputation.

Table 1. Root mean square errors (RMSE) for predictors of body mass index and computation times using different methods for imputation of missing values.

Imputation method	RMSE	Computation time
Median	5.43	-
Linear model	4.38	0.5 sec
Linear model with multiple imputation (MICE)	4.38	29 sec
K-nearest neighbour (VIM)	5.11	16 min
Random forest (CARET – RF)	4.40	4 h, 40 min
Neural network (CARET – NNET)	4.34	10 h, 12 min
Support vector machine (e1071)	4.49	13 min

3.3. Impact of Imputation on a Prediction Model

Finally, we compared the performance of a prediction model for the occurrence of delirium with three variations of data quality in BMI data. The model performance did not depend on the changes in BMI data and resulted in the same results for all three modelling scenarios. The accuracy for predicting the outcome of a delirium was 0.74, specificity 0.72 and sensitivity 0.88. Furthermore, the prediction outcome for those patients without a valid BMI did not differ between the scenarios with and without imputation.

Although there were no significant changes in the prediction performance, we observed a change in the variable importance when varying the BMI data quality. In the first modelling process the BMI data was excluded and therefore not presented as variable. In the second analysis, the variable for BMI was ranked as the 14th most important variable for prediction, whereas in the third scenario with cleansed and imputed data BMI was the second most important variable.

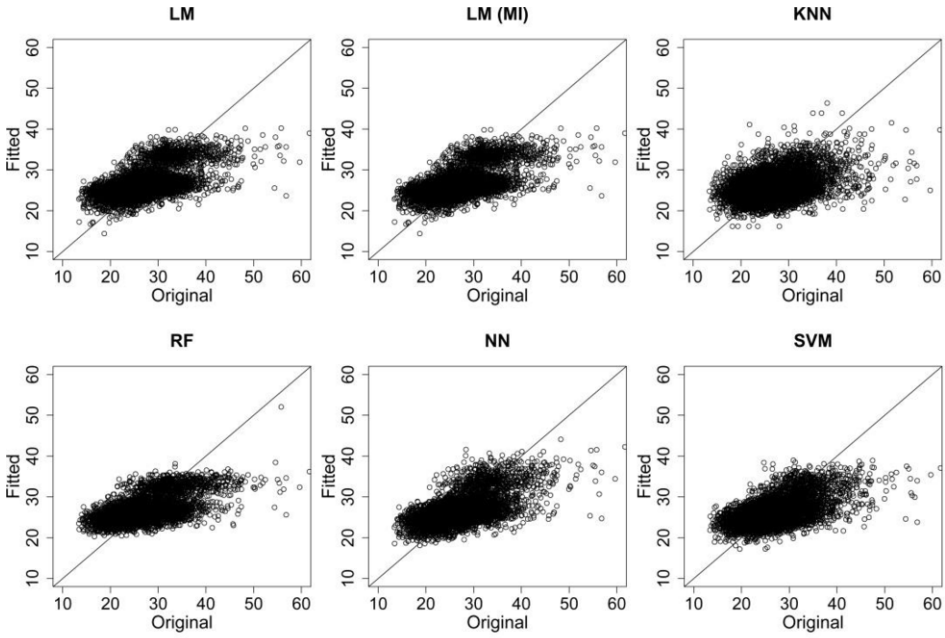


Figure 2. Scatterplots for original BMI values and fitted values for six imputation methods. LM – Linear model; LM (MI) – Linear model with multiple imputation; KNN – k-nearest neigbor; RF – Random forest; NN – Neural network; SVM – Support vector machine.

4. Discussion

In our study, we evaluated different imputation methods for BMI data and assessed their effect of cleansed and imputed data on a prediction model.

We cleansed the BMI data of 328,283 patients obtained from the EHR data of a Styrian HIS and compared imputation methods based on statistical methods and machine learning methods. RMSE was lowest for a NN predicting BMI in our test data set, but with the longest computation time. The fastest predictions with similar results were achieved by the linear regression models with simple and multiple imputation. Imputations based on ML methods varied in their quality of prediction: While RF and SVM resulted in RMSEs comparable to NN, the KNN method showed higher errors.

The use of prediction models in clinical practice requires high precision. Hence, there is a need for best performing prediction models. Despite the commonly used statement "garbage in, garbage out", the model performance in our simulation did not change with the variation of data quality. Even though BMI data was one of the most important predictors in our model and high in missing and incorrect values, the performance of the model was not affected. These results indicate that criticism of bad data quality should not be generalized for all scenarios of secondary use of data.

Although we did not observe a change in performance, the variable for BMI data gained in importance when cleansing and imputing methods were carried out. We assume this to be due to the information that was represented in imputed values: The imputation models included predictors that were afterwards used for risk prediction, e.g. diagnoses and lab data, resulting in a predictive model within a predictive model.

Several authors reported biases in BMI records. Even though it was not our objective to determine biases, further analysis of our data is needed to assess the importance of these and its impact on imputation methods.

A limitation of our study is the correctness of the BMI values that were used for imputation modelling. Although the data was checked for common errors and plausibility ranges were applied, some values might still be incorrect. We do not know how this fact influences the imputation modelling.

Finally, the imputed values need to be used with reservation. All compared imputation models had standard deviations of more than a BMI value of 4 kg/m^2 for predicted values. Considering descriptive statistics for BMI values, patients are likely to change quartiles due to imputation.

To sum up, our results showed that accuracy of imputation did not differ significantly between ML methods and statistical methods in general. If computation times need to be kept down, linear regression models may be preferable. Even though there was an effect of a variation in data quality on a prediction model considering the changes of variable importance, model performance did not change.

Acknowledgements

This work has been carried out with the K1 COMET Competence Centre CBmed, which is funded by the Federal Ministry of Transport, Innovation and Technology (BMVIT); the Federal Ministry of Science, Research and Economy (BMWFW); Land Steiermark (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the FFG.

KAGes and SAP provided significant resources, manpower and data as basis for research and innovation.

References

[1] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLOS ONE, 12(4), e0174944.
[2] Hao, S., Wang, Y., Jin, B., Shin, A. Y., Zhu, C., Huang, M., … Ling, X. B. (2015). Development, Validation and Deployment of a Real Time 30 Day Hospital Readmission Risk Assessment Tool in the Maine Healthcare Information Exchange. PLOS ONE, 10(10), e0140271.
[3] Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. A. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. Journal of the American Medical Informatics Association, 24(1), 198–208.
[4] Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. Cancer Informatics, 2, 59.
[5] Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artificial Intelligence in Medicine, 50(2), 105–115.
[6] Kontopantelis, E., Parisi, R., Springate, D. A., & Reeves, D. (2017). Longitudinal multiple imputation approaches for body mass index or other variables with very low individual-level variability: the mibmi command in Stata. BMC Research Notes, 10(1).
[7] Rea, S., Bailey, K. R., Pathak, J., & Haug, P. J. (2013). Bias in Recording of Body Mass Index Data in the Electronic Health Record. AMIA Summits on Translational Science Proceedings, 2013, 214–218.
[8] Mishra, G. D., & Dobson, A. J. (2004). Multiple imputation for body mass index: lessons from the Australian Longitudinal Study on Women's Health. Statistics in Medicine, 23(19), 3077–3087.
[9] De Laet, C., Kanis, J. A., Odén, A., Johanson, H., Johnell, O., Delmas, P., … Tenenhouse, A. (2005). Body mass index as a predictor of fracture risk: A meta-analysis. Osteoporosis International, 16(11), 1330–1338.
[10] Yang, H., Negishi, K., Otahal, P., & Marwick, T. H. (2015). Clinical prediction of incident heart failure risk: a systematic review and meta-analysis. Open Heart, 2(1), e000222.
[11] O'brien, T. E., Ray, J. G., & Chan, W.-S. (2003). Maternal body mass index and the risk of preeclampsia: a systematic overview. Epidemiology, 14(3), 368–374.
[12] Kramer, D., Veeranki, S., Hayn, D., Quehenberger, F., Leodolter, W., Jagsch, C., & Schreier, G. (2017). Development and Validation of a Multivariable Prediction Model for the Occurrence of Delirium in Hospitalized Gerontopsychiatry and Internal Medicine Patients. In Health Informatics Meets EHealth: Digital Insight–Information-Driven Health & Care. Proceedings of the 11th EHealth2017 Conference (Vol. 236, p. 32). IOS Press.
[13] Inouye, S. K. (2006). Delirium in Older Persons. New England Journal of Medicine, 354(11), 1157–1165.
[14] Freedman, D. S., Lawman, H. G., Skinner, A. C., McGuire, L. C., Allison, D. B., & Ogden, C. L. (2015). Validity of the WHO cutoffs for biologically implausible values of weight, height, and BMI in children and adolescents in NHANES from 1999 through 2012,. American Journal of Clinical Nutrition, 102(5), 1000–1006.
[15] Buuren, S. van, & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 1–68.
[16] Kuhn, M. (2012). caret: Classification and Regression Training.
[17] Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. New York: Springer-Verlag.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-124

Effect of Nursing Assessment on Predictive Delirium Models in Hospitalised Patients

Sai P. K. VEERANKI[a,1], Dieter HAYN[a], Diether KRAMER[b], Stefanie JAUK[c] and Günter SCHREIER[a]
[a]*AIT Austrian Institute of Technology, Graz, Austria*
[b]*Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria*
[c]*CBmed, Austria*

Abstract. Delirium is an acute neuropsychiatric syndrome which is common in elderly patients during their hospitalisation and is associated with an increased mortality and morbidity. Since delirium is a) often underdiagnosed and b) preventable if early signs are detected,igh expectations are set in delirium risk assessment during hospital admission. In our latest studies, we showed that delirium prediction using machine learning algorithms is possible based on the patients' health history. The aim of this study is to compare the influence of nursing assessment data on prediction models with clinical and demographic data. We approached the problem by a) comparing the performance of predictive models including nursing data with models based on clinical and demographic data only and b) analysing the feature importance of all available features. From our results we concluded that nursing assessment data can improve the performance of delirium prediction models better than demographic, laboratory, diagnosis, procedures, and previous transfers' data alone.

Keywords. Delirium, Predictive Analytics, Nursing Assessment

1. Introduction

Delirium is an acute neuropsychiatric syndrome that encompasses the symptoms of cognitive dysfunction including acute confusion, consciousness, disorientation, etc.[1-3]. It is a common clinical syndrome in elderly hospitalised patients [1,2]. It is often underdiagnosed [3,4] and can cause a cascade of events that would lead to a decrease in physical functionalities, longer hospital stays and can ultimately lead to death. However, it is assumed that delirium is preventable in 40% of the cases [5]. Therefore, early detection of individual risk factors might have a high relevance.

1.1. Screening and Assessment Instruments

According to the revisions in American psychiatric association's fifth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-V), the following criteria are used to diagnose delirium [6]:

A. Disturbance in attention (i.e., reduced ability to direct, focus, sustain, and shift attention) and awareness (reduced orientation to the environment).

[1] Corresponding Author: Sai Pavan Kumar Veeranki, AIT Austrian Institute of Technology, Reininghausstraße 13, 8020, Graz, E-Mail: sai.veeranki@ait.ac.at

B. The disturbance develops over a short period of time (usually hours to a few days), represents an acute change from baseline attention and awareness, and tends to fluctuate in severity during the course of a day.
C. An additional disturbance in cognition (e.g. memory deficit, disorientation, language, visuospatial ability, or perception).
D. The disturbances in Criteria A and C are not better explained by a pre-existing, established or evolving neurocognitive disorder and do not occur in the context of a severely reduced level of arousal such as coma.
E. There is evidence from the history, physical examination or laboratory findings that the disturbance is a direct physiological consequence of another medical condition, substance intoxication or withdrawal (i.e. due to a drug of abuse or to a medication), or exposure to a toxin, or is due to multiple aetiologies.

In clinical setups, physicians perform standardised assessment tools to understand the patient's cognitive impairments and the most commonly used tools are Mini Mental State Examination (MMSE) and Clock Drawing Test (CDT). The following assessment scales are based on the nurses´ observations during regular care: The Confusion Assessment Method (CAM) is the most widely used tool for identification of delirium both in clinics and research, the Memorial Delirium Assessment Scale (MDAS) is a clinically administered scale and it is used to rate the severity of delirium in cancer patients [7] and The Delirium Observation Screening Scale (DOSS) that was developed to facilitate early recognition of delirium according to DSM-IV [8].

1.2. Predicting Delirium Using Machine Learning Algorithms

Clinical delirium assessment as described above is time consuming and is often not performed. Therefore, there were already several approaches concerning application of machine learning techniques on delirium prediction that have been published in the last 10 years [9]. These approaches were based on evaluating different risk-stratification cohort scales that were mentioned above.

In one of the recent publications, Wassenaar et al. [9], made use of CAM-ICU for assessing the patient's delirium condition and applied a regression model on the data. The study cohort consisted of 2,914 patients in which 1,962 were included in the training and 952 in the validation dataset. They obtained an Area Under the Receiver Operating Characteristic (AUROC) of 70% with sensitivity and specificity of 62% and 67%, respectively, in the 0-1 day stay in ICU. These numbers increased to AUROC = 81% with a sensitivity of 78% and specificity 68% with six days of stay.

In an another study [10] with 397 patients that were hospitalised at internal medicine ward, a model has been developed based on rules derived from CAM. The model achieved an AUROC of 85% with sensitivity=80% and specificity=90%.

Many other studies on predicting delirium were published in the years before. An overview of risk-stratification models has been provided in a review by Newman et.al [11] in which the authors considered different risk factors and derived rules for predicting delirium [10,12-14]

Recently, our group published results from a delirium predictive model that was developed from a large cohort of hospitalised patients, which showed an AUROC of > 90 % [17]. When discussing these models and results with clinicians, we identified the interest to quantify the influence of nursing assessment data on the predictive model results – as compared to demographic and clinical data.

1.3. Objectives

The aim of this paper was to show the importance of the nursing assessment data in predicting delirium using machine learning algorithms in the hospitalised patients.

2. Methods

The routine health care data collected by the regional health care provider in Styria (Austria) Steiermärkische Krankenanstaltengesellschaft m.b.H (KAGes) were used in this study. For the past 16 years, the longitudinal Electronic Health Records (EHR) of almost 90% of the province's 1.2 million population were collected into the KAGes i.s.h.med based Hospital Information System (HIS).

2.1. Data Extraction

The EHRs of all patients were stored in the KAGes HIS and were made available in SAP HANA Platform (SAP, Walldorf, Germany), which supports extraction of relevant data and several data manipulation operations. Relevant data were extracted by applying the following inclusion and exclusion criteria for preparing a cohort and a control group.

2.2. Cohort Group

2.2.1. Inclusion Criteria

- Documented diagnosis of delirium (ICD-10 code F05.*) during hospitalisation in a KAGes hospital
- Date of admission between 2012 and 2017
- If delirium was diagnosed more than once for a patient, only the first admission with a delirium was included

2.2.2. Exclusion Criteria

- Delirium induced by alcohol, drugs and medications
- Additional diagnose of delirium (ICD-10 code F05.*) before the year 2012

Based on the inclusion criteria, we identified approximately 4,900 patients. After applying the exclusion criteria, we ended up with 3,800 patients.

2.3. Control Group

2.3.1. Inclusion Criteria

- No documented diagnosis of delirium (ICD-10 code F05.*)
- At least 1 hospitalisation in a KAGes internal medicine or geriatric department
- Date of admission between 2012 and 2017

From all patients who met the inclusion criteria of the control group, we randomly selected 25,000 patients and for each of these patients we randomly selected one of their hospitalised admissions to a KAGes hospital in the period of 2012-2017.

2.4. Data Extraction and Manipulation

Data were extracted from SAP HANA according to inclusion and exclusion criteria. Several data manipulations and summary statistics were applied through R software [15]. The features were classified into feature classes, which will be described in the following.

2.5. Relevant Data for the Cohort and Control Groups

Features of the following 6 features classes were derived from the data within the KAGes HIS for predictive modelling:

1. Demographic information
2. Laboratory results within the last thirty days from reference date
3. All diagnosis ever documented in the past
4. Health interventions (procedures) of the last two years from reference date
5. Previous transfers (admission, discharge, transfer, in-clinic / ambulatory, etc.)
6. General nursing assessment last recorded before reference date

An overview of the feature classes considered in our model is provided in Table 1. Reference date represents the date of admission into hospital for the selected case.

2.5.1. Data Processing and Modelling

Predictive Analytics Toolset for Health & Care (PATH) which was developed in MATLAB (The MathWorks, Natick, Massachusetts, United States of America) and was designed with special features for predictive modelling with healthcare data and has integrated all the standardised catalogues related to medicine especially for the German speaking countries such as International Classification of Diseases Version 10 (ICD-10), International Classification of Health Interventions (ICHI), International Classification of Procedures in Medicine (ICPM) or Logical Observation Identifiers Names and Codes (LOINC) etc.. Feature extraction, modelling and evaluation of models has been done using PATH. We selected random forest to analyse the effect of nursing assessment, since it had performed best in our previous analyses [17].

Table 1: Overview of the feature classes and respective features

Feature class	Number of features	Examples	Time period	
			From	To
Demographic	13	age, gender, education, district of residence	Any	Day of admission
Laboratory	70	e.g. CRP Value	30 d prior admission	Day of admission
Diagnosis	236	ICD-10 codes	All	Prior to admission
Procedures	87	MEL codes [16]	2 y prior admission	Prior to admission
Transfers	17	Time of admission and discharge, type of transfer, etc.	All	Prior to admission
Nursing Assessment	77	Hearing impaired, disoriented, non-German speaking, etc.	Any	Prior to admission

2.5.2. Feature Engineering

Diagnoses were binary expanded to one binary feature per ICD code and the number of all diseases per subchapter according to ICD-10 standard that were ever documented prior to the admission of the patient was calculated. Transfers prior to the current admission were used to extract relevant information such as number of ambulatory visits, number of hospitalisations, longest stay in hospital, duration between last discharge and current admission etc. Laboratory results that were selected from summary statistics were presented as ordinal values. Nursing assessment consists of a series of questions to assess the patient's cognitive status along with regular patient care information.

2.6. Evaluation

We have selected random forest to analyse the effect of nursing assessment, since it had performed best in our previous analyses [17]. We applied an ensemble method to obtain the probability of each patient to develop delirium during hospitalisation. We evaluated the models with different combinations of feature classes to analyse the influence of nursing assessment on the prediction performance. Feature importance was derived from random forests where all the available data were used based on the decrease in model performance when omitting single features. After applying a 10-fold-crossvalidation we obtained a confusion matrix, calculated sensitivity, specificity as well as the AUROC.

3. Results

With a feature set containing all the available data (demography, transfers, laboratory, procedures, diagnoses, and nursing), we have achieved an AUROC of 89.8%.

Figure 1 shows the improvement in AUROC as achieved when an additional feature class was added to models consisting of "reference feature sets" of increasing complexity (Fig. 1, from left to right): When added to an empty "reference feature set" (Figure 1, left), nursing data alone showed an AUROC of 81%, which was higher than the second-best feature class (demographic data, 79.4%). Even when added to any other "reference feature set", nursing feature class improved the results better than all other classes.

Figure 1. Effect of adding features of an additional feature class to different "reference feature set" on the AUROC in %. Content of the "reference feature set" is shown below the bar diagrams. Added feature classes are colour coded. The effect of each single feature class alone (starting from an empty "reference feature set") is shown on the left. Addition of additional feature classes to a "reference feature set" consisting of demographic data only, demographic + transfer data, etc. are shown in the subsequent diagrams.

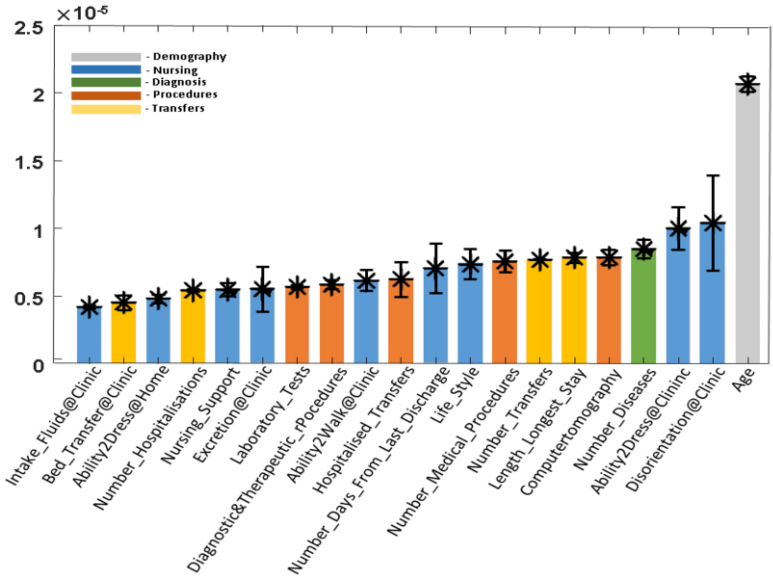


Figure 2. Twenty most important features based on feature importance obtained by the model developed using a random forest. The feature class of each feature is color coded. Mean ± Standard deviation of the feature importance as achieved for the different sub-models derived during 10-fold cross validation are illustrated.

Figure 2 shows the importance of the first twenty most important features as derived from the model that involved all the feature classes. While age was by far the most important feature for prediction, no other feature from the demography class was ranked among the best 20. However, nine out of these 20 features were part of the nursing dataset, including the second (disorientation at clinic) and third (ability to dress at clinic) important feature. Five features from the procedures feature class and four transfer features were ranked within the top 20 features.

4. Discussion

4.1. Interpretation of the Results

We have analysed which feature class contributes most to the results of our models when predicting delirium in hospitalised patients and how important the nursing features are as compared to the other data feature classes.

We found that – no matter whether starting from empty or from complex "reference feature sets", addition of nursing data improved the result better than data from all other feature classes. This indicates, that a) nursing data per se has a high potential in predicting delirium and b) nursing data provides additional and independent information, as compared to demographic and clinical data. These results were confirmed by our analyses of feature importance within our complete dataset, containing transfer, procedures, laboratory, diagnosis and nursing data.

These findings are remarkable, given the fact that – based on our experiences during implementation of the delirium prediction tool – physicians hardly rely on nursing assessments when assessing the risk of delirium for an individual patient.

The current related works, considered only delirium assessment information from the cases under observation for delirium prediction where as in our study we have shown that any past records of nursing assessment is helpful in improving delirium prediction performance along with complete EHR to predict the probability of developing delirium during the hospitalisation at the time of admission.

4.2. Limitations

Our model predicts the probability of delirium at hospital admission based on data from the past. Table 1 gives an overview over the time periods considered for each type of features. These restrictions had to be implemented in our retrospective model to avoid data leaks (inclusion of prospective data in the model). Therefore, patients who were admitted for the first time could only be assessed based on their demographic data and laboratory data from the day of admission (if available).

In a real-world scenario, additional data might be available, such as diagnoses and procedures documented on the day of admission but not prior to assessing the patient's delirium risk.

With a model containing all available features, we achieved an AUROC of 0.89. Based on this prediction, a threshold value can be selected to trigger action to prevent delirium. Selection of an optimal threshold depends strongly on the clinical scenario and interventions. However, in any case, a significant number of False Positive and False Negative predictions must be expected. Nevertheless, our model may help physicians to identify patients who may benefit from specific treatment to prevent delirium.

4.3. Outlook

The developed model is expected to go online as a pilot project on selected KAGes departments by the end of first quarter of 2018. The application will alert physicians with a yellow or red dot, when the predicted probability of delirium is above corresponding thresholds. In addition, the physicians will be provided with information of patient characteristics that contributed most in delivering such a result. Data will be collected prospectively, to assess the impact of the model on medical practice related to delirium. Based on these data, the model will be evaluated and refined, with the ultimate goal of providing a general model for predicting delirium in all the departments of KAGes network hospitals.

5. Conclusion

We conclude that – although often disregarded by physicians – nursing data are essential for predicting delirium at admission of patients to the hospital. Any past record of nursing data is worth to be considered for predicting delirium in hospitalised patients using machine learning algorithms.

Acknowledgement

This work is part of the IICCAB project (Innovative Use of Information for Clinical Care and Biomarker Research) within the K1 COMET Competence Centre CBmed (http://cbmed.at), funded by the Federal Ministry of Transport, Innovation and Technology (BMVIT); the Federal Ministry of Science, Research and Economy (BMWFW); Land Steiermark (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the FFG.

References

[1] T. G. Fong, S. R. Tulebaev, and S. K. Inouye, "Delirium in elderly adults: diagnosis, prevention and treatment," (in eng), Nat Rev Neurol, vol. 5, no. 4, pp. 210-20, Apr 2009.
[2] S. Wass, P. J. Webster, and B. R. Nair, "Delirium in the elderly: a review," (in eng), Oman Med J, vol. 23, no. 3, pp. 150-7, Jul 2008.
[3] Y. C. Chin, G. C. Koh, Y. K. Tay, C. H. Tan, and R. A. Merchant, "Underdiagnosis of delirium on admission and prediction of patients who will develop delirium during their inpatient stay: a pilot study," (in eng), Singapore Med J, vol. 57, no. 1, pp. 18-21, Jan 2016.
[4] C. Perelló Campaner, "[Delirium risk assessment in elderly hospitalized patients]," (in spa), Rev Esp Geriatr Gerontol, vol. 45, no. 5, pp. 285-90, 2010 Sep-Oct 2010.
[5] H. Zhang et al., "Strategies for prevention of postoperative delirium: a systematic review and meta-analysis of randomized trials," (in eng), Crit Care, vol. 17, no. 2, p. R47, Mar 2013.
[6] A. P. Association, "Diagnostic and Statistical Manual of Mental Disorders , Fifth Edition," ed, 2017.
[7] G. Sandeep and K. Natasha, "Assessment scales for delirium: A review," ed: World Journal of Psychiatry, 2012.
[8] A. P. Association, "Diagnostic and Statistical Manual of Mental Disorders, 4th Edition," ed, 2000.
[9] A. Wassenaar et al., "Multinational development and validation of an early prediction model for delirium in ICU patients," (in eng), Intensive Care Med, vol. 41, no. 6, pp. 1048-56, Jun 2015.
[10] J. A. Martinez et al., "Derivation and validation of a clinical prediction rule for delirium in patients admitted to a medical ward: an observational study," (in eng), BMJ Open, vol. 2, no. 5, 2012.
[11] M. W. Newman, L. C. O'Dwyer, and L. Rosenthal, "Predicting delirium: a review of risk-stratification models," (in eng), Gen Hosp Psychiatry, vol. 37, no. 5, pp. 408-13, 2015 Sep-Oct 2015.
[12] J. L. Rudolph et al., "Derivation and validation of a preoperative prediction rule for delirium after cardiac surgery," (in eng), Circulation, vol. 119, no. 2, pp. 229-36, Jan 2009.
[13] S. T. O'Keeffe and J. N. Lavan, "Predicting delirium in elderly patients: development and validation of a risk-stratification model," (in eng), Age Ageing, vol. 25, no. 4, pp. 317-21, Jul 1996.
[14] S. K. Inouye, Y. Zhang, R. N. Jones, D. K. Kiely, F. Yang, and E. R. Marcantonio, "Risk factors for delirium at discharge: development and validation of a predictive model," (in eng), Arch Intern Med, vol. 167, no. 13, pp. 1406-13, Jul 2007.
[15] S. W. Max Kuhn. Contributions from Jed Wing, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt., "caret: Classification and Regression Training. R Package Version 6. 0-73.," ed, 2016.
[16] "Krankenanstalten – selbstständige Ambulatorien – Leistungsorientierte Krankenanstaltenfinanzierung," S. Bundesministerium für Arbeit, Gesundheit und Konsumentenschutz, Ed., ed, 2017.
[17] D. Kramer et al., "Development and Validation of a Multivariable Prediction Model for the Occurrence of Delirium in Hospitalized Gerontopsychiatry and Internal Medicine Patients," presented at the Health informatics meets eHealth, Vienna, 2017.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-132

Development of a Protocol for Automated Glucose Measurement Transmission Used in Clinical Decision Support Systems Based on the Continua Design Guidelines

Markus MEYER[a], Klaus DONSA[a,1], Thomas TRUSKALLER[a], Matthias FROHNER[b], Birgit POHN[b], Alexander FELFERNIG[c], Frank SINNER[a] and Thomas PIEBER[a,d]
[a]*HEALTH - JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria*
[b]*University of Applied Sciences Technikum Wien, Vienna, Austria*
[c]*Graz University of Technology, Graz, Austria*
[d]*Medical University of Graz, Graz, Austria*

Abstract Background: A fast and accurate data transmission from glucose meter to clinical decision support systems (CDSSs) is crucial for the management of type 2 diabetes mellitus since almost all therapeutic interventions are derived from glucose measurements. Objectives: Aim was to develop a prototype of an automated glucose measurement transmission protocol based on the Continua Design Guidelines and to embed the protocol into a CDSS used by healthcare professionals. Methods: A literature and market research was performed to analyze the state-of-the-art and thereupon develop, integrate and validate an automated glucose measurement transmission protocol in an iterative process. Results: Findings from literature and market research guided towards the development of a standardized glucose measurement transmission protocol using a middleware. The interface description to communicate with the glucose meter was illustrated and embedded into a CDSS. Conclusion: A prototype of an interoperable transmission of glucose measurements was developed and implemented in a CDSS presenting a promising way to reduce medication errors and improve user satisfaction.

Keywords. Mobile Health, Standardization, Clinical Decision Support Systems, Type 2 Diabetes Mellitus, Medication Errors.

1. Introduction

425 million people are suffering from diabetes mellitus worldwide and 58 million people in Europe with around 90% of type 2 diabetes mellitus (T2DM) according to recent estimates of the International Diabetes Federation (IDF) [1]. Glucose measurements and medication administration are important components of T2DM therapy.

Each patient needs a different amount of insulin which depends on many internal and external factors. General Practitioners (GPs) need comprehensive diabetes knowledge and experience to set a personalized insulin dosage for a patient by analyzing the logged glucose measurements. Since it is a time-consuming task to identify the cause of every Hypo- and Hyperglycemia, GPs can usually only focus on the most recent ones. A clinical decision support system (CDSS) can analyze large datasets of measurements

[1] Corresponding Author: Klaus Donsa, HEALTH - JOANNEUM RESEARCH Forschungsgesellschaft mbH, Neue Stiftingtalstraße 2, 8010 Graz, E-Mail: healthca@joanneum.at.

in a short time and can therefore assist GPs by suggesting a personalized medication dosage derived from all logged glucose measurements.

GlucoTab® is a CDSS which is already used in hospitals for the therapy of T2DM. Medication dosage suggestions are derived from glucose measurements as well as personalized therapy settings and follow a rule-based algorithm. The algorithm in place frequently updates the therapy settings based on new input data to fit the personal needs of the patient. GlucoTab® is currently operated by healthcare professionals (HCPs), however with the limitation that the glucose measurements are transferred manually into the CDSS.

The most important variable of a rule-based algorithm to calculate insulin dosages is the blood glucose level. A previously performed study [2] showed an error rate of 5% during manual transfer of measured glucose concentrations from the glucose meter in a paper-based workflow and an error rate of 4% in a computerized workflow. The study further revealed an increased probability of a hypoglycemic event following an insulin dosing error (odds ratio 3.1). Severe hypoglycemia is an indicator for poor patient outcomes and higher mortality risk and should therefore be avoided [3]. Preventing errors by the transfer of glucose measurements with an automated transmission protocol can therefore reduce the mortality rate.

Almost all therapeutic actions are derived from the measured glucose concentrations and therefore an accurate transmission of glucose measurements to other health-related devices such as CDSSs is desirable. Aim was the development of a prototype of an automated glucose measurement transmission protocol based on the Continua Design Guidelines. The protocol will be embedded into a mobile CDSS which will be used by healthcare professionals for the treatment of T2DM patients in the home care setting.

2. Methods

We performed a structured literature and market research on measurement transmission protocols to retrieve state-of-the-art implementations and to develop a prototype of an automated glucose measurement transmission protocol.

2.1. Literature and market research

The query “IEEE 11073 (Medical OR Health) Device” was used to search for publications about protocols, systems and devices which use the personal health device communication standard as defined by ISO/IEEE 11073. The query is applied to IEEE Xplore, ACM and PubMed Digital Library with a total of 99 distinct results. Based on title and abstract, 57 publications were identified as non-relevant for the research because they did not comprise the topic of a personal health device. Titles and abstracts of the 42 remaining relevant publications were examined and rated on a scale from 1 to 10 according to their relevance. This resulted in 21 relevant papers with a ranking of 5 or higher.

We used the google search engine and google play store to identify the state-of-the-art of glucose meters and smartphone applications related to automated measurement transmissions. Glucose meters are categorized by their data transport type (Bluetooth/ZigBee/USB/NFC) and whether they are Continua certified or not.

Table 1. Glucose meters, data transport and standardization status.

Glucose Meter	Data Transport	Continua Certified
Accu-Chek® Guide	Bluetooth/USB	Yes
Accu-Chek® Instant	Bluetooth/USB	Yes
Contour® Next/Plus ONE	Bluetooth/USB	Yes
FORA® D40	Bluetooth/USB	Yes
Accu-Chek® Active	USB	Yes
Accu-Chek® Mobile	USB	Yes
Accu-Chek® Aviva/Performa Insight	USB	Yes
Abbott FreeStyle Libre	NFC	No
Accu-Chek® Aviva/Performa Connect	Bluetooth/USB	No
AgaMatrix Jazz Wireless 2	Bluetooth	No
Beurer GL 50 evo	Bluetooth/USB	No
BodyTel® GlucoTel	Bluetooth	No
Dexcom G5	Bluetooth	No
FORA® TN'G / TN'G Voice	Bluetooth	No
GlucoMen Areo / Areo 2K	Bluetooth/NFC/USB	No
MediTouch® 2 connect	Bluetooth/USB	No
Medtronic Enlite® Sensor	Bluetooth	No
OneTouch Verio Flex®	Bluetooth/USB	No

Applications are categorized by features like reminders and bolus calculators, as well as whether they are using a standardized or a proprietary communication protocol.

2.2. Development of a prototype of an automated glucose measurement transmission protocol

Results from the literature and market research were the basis for the development of an automated glucose measurement transmission protocol from glucose meters to CDSSs. Literature research highlighted the benefits of standardized communication protocols which confirmed the development following the Continua Design Guidelines. Market research revealed a lack of glucose meters using a standardized communication protocol and was resolved by a middleware for the communication with glucose meters which can be extended to translate non-standardized messages into standardized messages.

The system was designed to transfer the measurements from an Accu-Chek® Guide glucose meter via Bluetooth Low Energy to a middleware running on Android. The middleware shall then provide an interface to other applications and allow them to receive and read measured glucose concentrations. The CDSS GlucoTab® implements the interface to the middleware and is thereby able to make therapy decisions from the sensor readings in real-time.

3. Results

Results from literature and market research were used as basis for development of a standardized glucose measurement transmission protocol using an extendable middleware to allow the support of a broad range of glucose meters.

3.1. Literature and market research

Market research revealed that at least 18 glucose meters which can transfer a measured glucose value to another device for further examination exist. However, only 7 out of the

18 listed devices use the standardized ISO/IEEE 11073-20601 PHD exchange protocol [4] and ISO/IEEE 11073-10417 glucose meter device specialization [5], as suggested by the Continua Design Guidelines (Table 1). Moreover only 4 of them support a wireless Bluetooth communication which increases usability in a mobile healthcare setting.

Market research on Android applications which are used to receive measurement values from a glucose meter showed a lack of standardization likewise in software applications and in personal health devices. Four out of 12 listed applications support the standardized protocol as defined by the Continua Design Guidelines (Table 2). Market research revealed further that most applications can only receive, store and visualize the measurement values, but they neither help the user with the medication dosage calculation nor with reminders for glucose measurements or medication administration.

Accu-Chek® Connect and mySugr provide a bolus calculator. The application thereby suggests an insulin dosage for a meal, based on carbohydrates, measured blood glucose and some personalized therapy settings, such as the amount of insulin needed per gram carbohydrate. GlucoTab® takes this approach one step further and manages the entire diabetes therapy. The application tells the user when and how often a glucose measurement should be performed and calculates an appropriate medication dosage several times per day.

3.2. Development of a prototype of an automated glucose measurement transmission protocol

The developed glucose measurement transmission protocol was designed and implemented according to the Continua Design Guidelines [6]. These guidelines define, beside other things, what has to be considered when implementing a Bluetooth Low Energy interface between a personal health device (the Accu-Chek® glucose meter) and a personal health gateway (the tablet running GlucoTab®). This specification is strongly aligned with the specification from the Bluetooth Special Interest Group. An end use application of the protocol was embedded into GlucoTab® (Figure 1).

GlucoTab® is installed together with a middleware on the used tablet. The middleware was implemented as a background service (not having an own graphical user interface) and provides the needed functionalities to connect and communicate with the used Accu-Chek® glucose meter on the one and GlucoTab® on the other hand.

Table 2. Feature, protocol and user comparison of diabetes related android applications.

Application	Logbook	Medication Calculator	Therapy Adjustment	Reminder	Continua Protocol	User
GlucoTab®	Yes	Yes	Yes	Yes	Yes	HCP
mySugr	Yes	Yes	No	Yes	Yes	P
Accu-Chek® Connect	Yes	Yes	No	No	Yes	P
Contour® Diabetes	Yes	No	No	Yes	Yes	P
AgaMatrix Diabetes Manager	Yes	No	No	Yes	No	P
Beurer HealthManager	Yes	No	No	No	No	P
BodyTel Blutzucker	Yes	No	No	No	No	P
Glucolog Lite/Mobile	Yes	No	No	No	No	P
iFORA Diabetes Manager	Yes	No	No	No	No	P
LibreLink	Yes	No	No	No	No	P
OneTouch Reveal®	Yes	No	No	No	No	P
VitaDock+	Yes	No	No	No	No	P

Figure 1. Transmission from Accu-Chek® Guide glucose meter to GlucoTab® CDSS.

Communication between middleware and GlucoTab® were defined through Android Interface Definition Language (AIDL). For the information that is exchanged based on the data objects (Figure 2) stated in the interfaces' definition, further efforts towards standardization were made by using the terminology defined in ISO/IEEE 11073-10417[5], i.e. the terms and/or code as stated in this standard were used to identify the measured parameters and meta data. Following this standard, information like measuring the glucose concentration based on an capillary whole-blood sample or meta information like that measurement has been taken pre-prandial was coded as the integer values 23112 and 29260 respectively.

The middleware, acting as a glucose meter manager, implements an IGlucoseMeterManager interface (Figure 3) and handles communication with the glucose meter as well as with GlucoTab®. Android's build-in platform support for BLE, which is available since Android 4.3, is used to read the services provided by a remote BLE device.

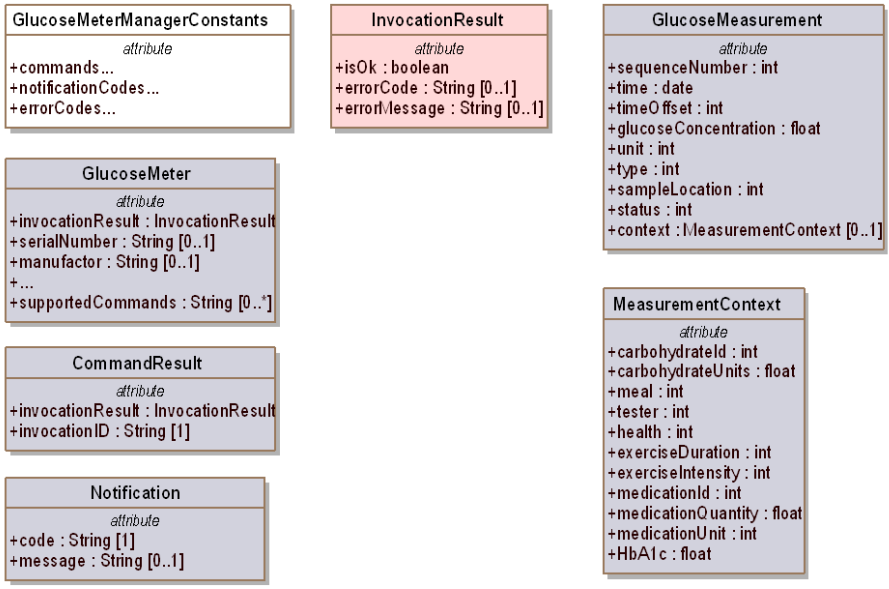


Figure 2. Data objects exchanged by middleware and GlucoTab®.

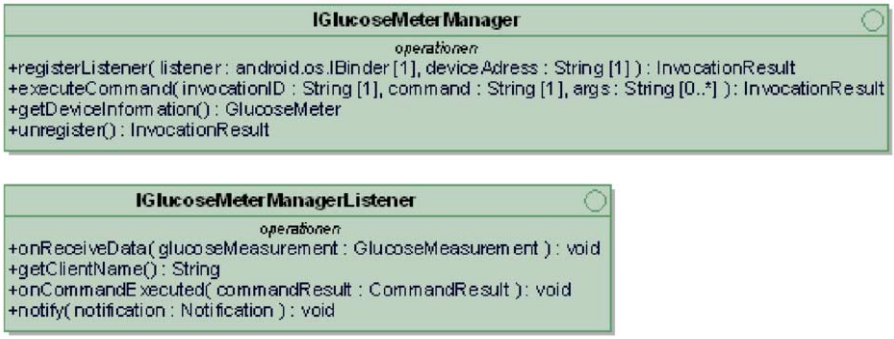


Figure 3. AIDL interface definition between middleware and GlucoTab®.

GlucoTab® implements the IGlucoseMeterManagerListener interface and registers as listener at the middleware by calling the registerListener method. This method holds the unique MAC address of the Bluetooth device to which the connection should be established. After the middleware has established the connection the "Record Access Control Point" service characteristic is used to query actively for stored values. This method also enables the listener to delete stored measurement values on the glucose meter. Methods of the IGlucoseMeterManager interface return a lightweight InvocationResult object to identify errors and supply error details with an error code and a short error message.

The GlucoseMeter class returned by the getDeviceInformation method contains general information about a connected glucose meter and additionally provides information about supported commands by the health device. GlucoseMeasurement is used to provide general information of a glucose measurement, such as the date and time and the glucose concentration. MeasurementContext can be used to give further details about a glucose measurement, such as exercise, meal and medication information related to the measurement when supported by the used blood glucose meter.

4. Discussion

Literature and market research substantiate the need for standardization in communication protocols used by personal health devices. As mentioned in [7-15], a standardized communication protocol for personal health devices enables a seamless plug and play compatibility of various sensors from different manufacturers. However, as mentioned by [8,11,13,15], manufacturers use their own software and communication protocols, building proprietary solutions that can only work alone or inside a single-vendor system. Proprietary protocol solutions eliminate the communication between devices of different manufacturers, leading to an interoperability problem.

One reason for the lack of standardization is the gap between current regulations as well as health policies of medical devices and stakeholders manufacturing the personal health technology. Insufficient or over regulation of health standards can both significantly delay the market adoption [12]. A further reason is the vast amount of pages in standard documents that need to be observed [13] and the overly complex design of the protocols [16]. Nevertheless, advantages of a standardized protocol are the interoperability of devices from different manufacturers, lower healthcare costs and a

better patient treatment [11-13]. Moreover, there are tools and frameworks provided to help overcome the difficulties of implementing a standardized communication protocol [13,17].

This article presents a standardized implementation of an automated glucose measurement transmission protocol from glucose meters to the CDSS GlucoTab®. By following the Continua Design Guidelines, future Continua certified glucose meters will be able to communicate out of the box with the presented implementation. Beside the implemented feature to query data using the "Record Access Control Point" characteristic another approach would be to have the data transferred automatically using the indication service, as described by the "Glucose Measurements" service characteristic specification [18]. Concerning the different possibilities how to acquire the data from the Bluetooth device, initiated by the device versus initiated by the middleware or GlucoTab®, at this point in time the decision was made for the latter.

At the first glance an automatic transfer of the data seems more desirable, but for the workflow of HCPs it seems more usable when they can actively pull the values. Therefore the "Record Access Control Point" service characteristic has been implemented and used to get and to delete data from the Bluetooth device. However, another possible way for the same user experience would be that the middleware stores and forwards the data received from the glucose meter. This approach has the disadvantage that data might end up in the middleware and might not be requested from GlucoTab®.

By embedding the automated glucose measurement transmission protocol into GlucoTab® and thereby eliminating the risk of miscopied or misread glucose measurement values, the treatment of T2DM patients by healthcare professionals was further improved. After measuring the glucose concentration of a patient, the value will automatically be transmitted to GlucoTab®. The clinical decision support system can immediately feed the measured value into an algorithm to calculate an adapted medication dosage based on the current glucose level.

Further research including a summative usability assessment of the automated transmission protocol using GlucoTab® in a working environment with healthcare professionals is already planned.

In conclusion a prototype of an automated glucose measurement transmission protocol was developed and embedded into GlucoTab®. Since the glucose measurement is the source of the algorithm behind GlucoTab® a reliable and automated transmission of this data helps to reduce medication errors and to assist HCPs on routine tasks.

References

[1] I. D. Federation, IDF Diabetes Atlas, Eighth edition 2017, 8th ed. 2017.
[2] K. Donsa et al., "Impact of errors in paper-based and computerized diabetes management with decision support for hospitalized patients with type 2 diabetes. A post-hoc analysis of a before and after study," Int. J. Med. Inform., vol. published, no. http://dx.doi.org/10.1016/j.ijmedinf.2016.03.007, 2016.
[3] A. K. Lee et al., "The Association of Severe Hypoglycemia With Incident Cardiovascular Events and Mortality in Adults With Type 2 Diabetes.," Diabetes Care, vol. 41, no. 1, pp. 104–111, Jan. 2018.
[4] IEEE 11073 Committee, IEEE 20601 - Health informatics—Personal health device communication: Application profile — Optimized Exchange Protocol, vol. 2014. 2014.
[5] I. 11073TM S. Committee, "IEEE 11073-10417-2015 - Health informatics -- Personal health device communication Part 10417: Device Specialization -- Glucose Meter." p. 78, 2015.
[6] Personal Connected Health Alliance, "H.811 Personal Health Devices Interface Design Guidelines," p. 114, 2016.

[7] H. S. Park, H. Cho, and H. S. Kim, "Development of cell phone application for blood glucose self-monitoring based on ISO/IEEE 11073 and HL7 CCD," Healthc. Inform. Res., vol. 21, no. 2, pp. 83–94, 2015.
[8] J. H. Lim, C. Park, and S. J. Park, "Home healthcare settop-box for senior chronic care using ISO/IEEE 11073 PHD standard," 2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC'10, pp. 216–219, 2010.
[9] M. Clarke, S. Member, J. De Folter, C. Palmer, and V. Verma, "Building Point of Care Health Technologies on the," pp. 16–18, 2013.
[10] A. Aragüés et al., "Trends and challenges of the emerging technologies toward interoperability and standardization in e-health communications," IEEE Commun. Mag., vol. 49, no. 11, pp. 182–188, 2011.
[11] I. Martińez et al., "Standard-based middleware platform for medical sensor networks and u-health," Proc. - Int. Conf. Comput. Commun. Networks, ICCCN, pp. 714–719, 2008.
[12] D. Zhong, M. J. Kirwan, and X. Duan, "Regulatory barriers blocking standardization of interoperability," J. Med. Internet Res., vol. 15, no. 7, 2013.
[13] J. H. Lim, C. Park, S. J. Park, and K. C. Lee, "ISO/IEEE 11073 PHD message generation toolkit to standardize healthcare device," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, pp. 1161–1164, 2011.
[14] H. L. Li, Z. Bin Duan, J. Z. Cui, and Z. W. Chen, "A design of general medical data adapter based on ISO/IEEE 11073 standards," 2015 12th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2015, pp. 404–407, 2016.
[15] J. C. Nam, W. K. Seo, J. S. Bae, and Y. Z. Cho, "Design and development of a u-Health system based on the ISO/IEEE 11073 PHD standards," 17th Asia-Pacific Conf. Commun. APCC 2011, no. October, pp. 789–793, 2011.
[16] M. Clarke, J. de Folter, V. Verma, and H. Gokalp, "Interoperable End-to-End Remote Patient Monitoring Platform based on IEEE 11073 PHD and ZigBee Health Care Profile," IEEE Trans. Biomed. Eng., vol. 1, no. c, 2017.
[17] L. Yu, Y. Lei, R. N. Kacker, D. R. Kuhn, R. D. Sriram, and K. Brady, "A general conformance testing framework for IEEE 11073 PHD's communication model," Proc. 6th Int. Conf. PErvasive Technol. Relat. to Assist. Environ. - PETRA '13, pp. 1–8, 2013.
[18] B. SIG, "Bluetooth GATT Characteristic: Glucose Measurement," 2012. [Online]. Available: https://www.bluetooth.com/specifications/gatt/viewer?attributeXmlFile=org.bluetooth.characteristic.glucose_measurement.xml. [Accessed: 25-Jan-2018].

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-140

Identifying and Validating Requirements of a Mobile-Based Self-Management System for People Living with HIV

Esmaeil MEHRAEEN[a], Reza SAFDARI[a,1], SeyedAhmad SEYEDALINAGHI[b], Niloofar MOHAMMADZADEH[a], and Goli ARJI[a]
[a] *Department Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran.*
[b] *Iranian Research Center for HIV/AIDS, Iranian Institute for Reduction of High Risk Behaviors, Tehran University of Medical Sciences, Tehran, Iran.*

Abstract. Background: Due to the widespread use of mobile technology and the low cost of this technology, implementing a mobile-based self-management system can lead to adherence to the medication regimens and promotion of the health of people living with HIV (PLWH). We aimed to identify requirements of a mobile-based self-management system, and validate them from the perspective of infectious diseases specialists. Method: This is a mixed-methods study that carried out in two main phases. In the first phase, we identified requirements of a mobile-based self-management system for PLWH. In the second phase, identified requirements were validated using a researcher made questionnaire. The statistical population was infectious diseases specialists affiliated to Tehran University of Medical Sciences. The collected data were analyzed using SPSS statistical software (version 19), and descriptive statistics. Results: By full-text review of selected studies, we determined requirements of a mobile-based self-management system in four categories: demographic, clinical, strategically and technical capabilities. According to the findings, 6 data elements for demographic category, 11 data elements for clinical category, 10 items for self-management strategies, and 11 features for technical capabilities were selected. Conclusion: Using the identified preferences, it is possible to design and implement a mobile-based self-management system for HIV-positive people. Developing a mobile-based self-management system is expected to progress the skills of self-management PLWH, improve of medication regimen adherence, and facilitate communication with healthcare providers.

Keywords. Self-management, Mobile, Data element, HIV/AIDS

1. Introduction

HIV-positive people require more support for the management of their condition, including making physical, psychological, and social adjustments because of the chronic condition of the disease [1,2]. Improving the condition of people living with HIV (PLWH) is dependent not only on healthcare services but also on social support and the provision of educational information in several areas such as how to make safe sex behaviors, antiretroviral therapy (ART) and adherence to medication regimens [3-5]. Among these factors, ART is more important that depends on the timely take of prescribed medications, diet and exercise compliance [6,7]. Self-management can

[1] Corresponding author: Reza SAFDARI, Tehran University of Medical Sciences, Qhods Square, Post Code: 7134845794, Tehran, Iran. E-mail: rsafdari@tums.ac.ir

provide effective solutions to motivation and contribution of patients for continuing in the effective therapeutic goals [8].

Self-management is a set of deliberated, learned, and purposeful activities that carried out by persons with chronic conditions targets reducing disease progression, managing symptoms and the prevention of disabilities [9-11]. Self-management is not a substitute for professional and organizational services, but it is a supplement of healthcare services and as a supportive way to implement specific strategies [12]. In recent years, providing self-management services through mobile technology has been introduced as a new approach to improve health care services and cost savings for HIV-positive people [13,14]. The application of mobile technology in the healthcare industry is widely regarded as a new way of supporting public health services [15-17] and can be used by PLWH to support self-management strategies [18]. According to the widespread use of mobile technology and the low cost of this application, developing a mobile-based self-management system can improve ART, physical, psychosocial, health knowledge, and behavioral outcomes of PLWH [19].

Mobile-based self-management systems with common requirements and technical capabilities to improving HIV-care are expressively required in limited resources countries that are bearing the effect of the HIV epidemic [20]. Unfortunately, in spite of the alarming occurrence and consequences of HIV infection, a self-management system with effective data elements and features are just now being deployed in a few countries [21]. In the same way, there have been limited strategies in developing countries to management of HIV chronic condition [22,23]. As a result, without self-management strategies and technical capabilities to adherence at the HIV care intervention, it is impossible to take an experimental approach to improve HIV care services [24]. However, in Iran, as a developing country, there has not been a mobile-based self-management system for HIV-positive people with the comprehensive data elements and technical capabilities to date [25].The purpose of this study was to 1) identifying requirements of a mobile-based self-management system for PLWH, and 2) validating of identified requirements based on infectious diseases specialists' attitudes.

2. Methods

This study was a descriptive cross-sectional one that carried out in two main phases in 2017. In the first phase, we searched documents to identify of requirements and common elements of a mobile-based self-management system for people living with HIV. The combination keywords used to search for resources included in; self-care, self-management, self-monitoring, data elements, minimum data set, requirements, mobile application, smart phone, mobile health, strategy, and HIV/AIDS that have been searched in PubMed, Web of Science, Up To Date, Science direct, Scopus, and Ovid. Studies were included if they reported on common data elements, HIV-care strategies, and technical capabilities of mobile-based HIV/AIDS self-management system. Moreover, full text contents research papers and review articles published between 2000 and 2017 and in English language were selected. We excluded newspapers, abstracts, editorial letters, and reports.

In the second phase, using the identified requirements from the review of related articles [1, 3-15, 18-21, 25, 29, 49], a questionnaire was developed to validate identified requirements. The questionnaire consisted of 4 categories and 47 questions (demographic data elements: 8 questions, clinical data elements: 16 questions, HIV-

Table 1. Search result and final reviewed articles

Elements	Total Reference Retrieved	Total Duplicate References	Total Excluded References	Final Analyzed Articles
Demographic	40	16	20	4
Clinical	57	31	20	6
Technical	70	35	30	5
Strategically	109	55	41	13
Total	276	137	111	28

management strategies: 10 questions, and technical features: 12 questions). Reliability of the questionnaire was calculated 0.87 by Cronbach's alpha and the validity was measured by seven specialists of infectious diseases and health information management experts. The statistical populations were all infectious diseases specialists working at Tehran University of Medical Sciences (N=23). Twenty-one individuals out of 23 statistical population completed the questionnaire. The data were analyzed using descriptive statistics and frequency distribution reports. In this way, the questionnaire options were scored from 1 to 5 (completely agree=5, agree=4, no idea=3, disagree=2, and completely disagree=1). Each of the identified requirements was considered as preference element that had obtained at least a mean of 2.5 or more.

3. Results

Using applied research strategies, 276 references were retrieved and finally, 28 related articles that published between 2000 and 2017 were reviewed. Table 1 shows the final analyzed articles, and the resources that did not meet inclusion criteria and excluded from this study. By full text review of selected articles we determined the requirements for a mobile-based self-management system in four categories: demographic (8 data elements), clinical (16 data elements), technical capabilities (12 features) and HIV self-management strategies (10 items).

According to the infectious diseases specialists' attitudes, "height" and "body mass index (BMI)" were not selected as required demographic data elements and only six data elements were selected for this category. In clinical category, mean score for "rash", "sore throat", "fatigue & tiredness", "vital signs", "associated diseases", and "current medication" was less than 2.5 and therefore, 11 data elements were selected for clinical category. According to the findings, all of the identified technical requirements were selected by infectious diseases specialists, except "data collection". Moreover, the mean score for all identified self-management strategies was more than 2.5, and therefore, all of them were selected. The selected data elements for each of categories and mean scores assigned to them by the specialists are shown in Table 2.

As shown in Table 2; 38 data elements and features were selected for different identified categories as requirements of a mobile-based self-management system for PLWH.

4. Discussion

People with a chronic condition have a central role in managing their conditions. These individuals can play an important role in improving their health status by benefiting from standard self-management programs [26-28]. A mobile-based self-management system

Table 2. The mean of given values and selected elements, features and strategies

Demographic			Technical capabilities		
Data element	**Mean**		**Feature**	**Mean**	
Gender	4.6	√	Content of text messages	3.3	√
Age	4.4	√	Educational messages	4.9	√
Marital status	4.5	√	Motivational messages	4.4	√
Education level	2.9	√	Drug taking reminder	5	√
Occupation	3.3	√	Appointment reminder	3.9	√
Height	2.3		Exercise reminder	4.6	√
Weight	4.2	√	Diet reminder	4.9	√
Body Mass Index (BMI)	2.4		Instructions	3.3	√
Clinical			Data collection	1.7	
Fever	4.9	√	Internet access	4.2	√
Chills	4.5	√	Being user-friendly	4.5	√
Night sweats	2.9	√	Security requirements	4.9	√
Weight loss	3.3	√	**Self-management strategies**		
Fatigue & Tiredness	1.8		Safe sexual behavior	3.9	√
Sore throat	2.3		Education	4.1	√
Mouth ulcers	2.6	√	Communication	2.8	√
Muscle pain	4.2	√	Motivational messages	4.3	√
Skin problems	3.3	√	Physical activity improvement	2.9	√
Diarrhea	4.1	√	Nutrition regimen	3	√
Pneumonia	4.6	√	Symptom management	4.2	√
Swollen lymph nodes	4.7	√	ART and medication adherence	5	√
Neurological disorders	3	√	Attend appointments	3.8	√
Current medication	2.3		Enhance quality of life	4	√
Associated diseases	2				
Vital signs	2.3				

can serve as a healthcare complement and includes areas such as maintaining health promotion, lifestyle modification, medication adherence, symptom assessment, disease management and rehabilitation [29-31]. In this mixed method study, we aimed to identify requirements of a mobile-based HIV self-management system, and validate them from the perspective of infectious diseases specialists.

The results of this study showed that demographic, clinical, technical, and strategies categories are essential for a mobile-based self-management system for HIV-positive people. Navato et al, in a 2015 developmental study, identified requirements of a mobile messaging system for tracking the care of PLWH and Tuberculosis (TB) in six categories: data acquisition requirements, telecommunications cost, privacy and data security, text message content, communication, and system scalability. The findings of this study showed that using this system could improve self-management and self-care skills of patients and strengthen the relationship between patients and health care providers [32]. Because of the importance of therapeutic objectives of HIV condition, existence of a mobile-based system that is focused on providing self-management services can be an important achievement for healthcare organizations [33]. Organized clinical data elements are a primary requirement for a mobile-based self-management system for information management and providing efficient clinical procedures in healthcare organizations [34]. Furthermore, complete registration of patients' demographic information in a self-management system will help them better identify and manage their prescribed treatment [24]. Demographic data elements of PLWH in a comprehensive HIV/AIDS information management system should be documented in

order to better understand the epidemic of the disease, and to focus on sub-populations and management of treatment [25].

According to the findings, technical capabilities of a mobile-based self-management system for PLWH needs features such as; educational and motivational messages, content of the text messages, drug taking reminder, and security requirements. Technical capabilities such as adherence to medication, drug taking reminder and data security are the important aspects of a mobile-based self-management system [35-38].This important finding was highlighted in most similar studies that, technical capabilities of a mobile-based self-management system can be very helpful by giving reminders and engaging patient in therapeutic activities [39]. For example, a 2016 clinical trial by Garofalo et al., illustrated that a text message-based system in addition to reminders of medication adherence and attend appointments, can also facilitate communication with healthcare providers [40].

Based on the findings of this study, 10 mobile-based self-management strategies were identified and selected for HIV-positive people. Access to health care services and the motivation of HIV-positive people for self-management are two important factors in improving their health status [41-43]. ART is one of the most important self-management strategies and clinical prescriptions for the treatment of HIV, especially in developing countries where the number of PLWH is rising [44,45]. In similar studies, it has been recommended that PLWH should participate in ART strategy and collaborate with healthcare providers to succeed in their clinical treatment [46-50]. Self-management strategies, in addition to ART, should also emphasize on adherence to medication [51]. Adherence to compound medication regimens often leads to a significant aspect of chronic condition management. PLWH are required to start medication regimens that demand a high degree of adherence and attend medical appointments [52]. Furthermore, regarding HIV transmission ways and the social stigma of this disease, it is important to provide educational information on safe sex approaches and how to prevent the transmission of the virus as well as to strengthen the morale of PLWH [53-55].

5. Conclusion

In this study, we determined four requirements categories (demographic data elements, clinical data elements, technical capabilities and HIV-care strategies) of mobile-based self-management system for people living with HIV. Using these requirements, it is possible to design and implement a mobile-based self-management system for PLWH. Implementation of this mobile-based system can provide a timely reminder to improve medication adherence, promote the self-management skills, and safe-sex negotiation for HIV-positive people. Complementary studies to address the preferences of an intelligent mobile-based self-management system for PLWH can be an appropriate course for future researches on this topic.

Acknowledgments

This manuscript was based on some results of the Ph.D. dissertation with titled "Design, Develop and Assessment a Mobile and Web-based Intelligent Self-Management System for HIV Positive Persons" on health information management at Tehran University of

Medical Sciences. The authors extend thanks to infectious diseases specialists working at Tehran University of Medical Sciences.

References

[1] van Olmen J, Schellevis F, Van Damme W, Kegels G, Rasschaert F. Management of chronic diseases in sub-Saharan Africa: cross-fertilisation between HIV/AIDS and diabetes care. Journal of tropical medicine. 2012. doi:10.1155/2012/349312

[2] Khalili H, Rohani R, Seyedalinaghi S, Hajiabdolbaghi M, Dashti-Khavidaki S, Hajhosssein Talasaz A. Adherence to antiretroviral therapy among Iranian HIV/AIDS patients. Current clinical pharmacology. 2012;7(2):111-5.

[3] Woods SP, Iudicello JE, Morgan EE, Cameron MV, Doyle KL, Smith TV, Cushman C, HIV Neurobehavioral Research Program (HNRP) Group, Grant I, Atkinson JH, Ellis RJ. Health-related everyday functioning in the internet age: HIV-associated neurocognitive disorders disrupt online pharmacy and health chart navigation skills. Archives of Clinical Neuropsychology. 2016 Jan 6;31(2):176-85.

[4] Stoller EP, Webster NJ, Blixen CE, McCormick RA, Perzynski AT, Kanuch SW, Dawson NV. Lay management of chronic disease: a qualitative study of living with hepatitis C infection. Am J Health Behav, 2009;33(4):376-90.

[5] Iribarren S, Siegel K, Hirshfield S, Olender S, Voss J, Krongold J, Luft H, Schnall R. Self-Management Strategies for Coping with Adverse Symptoms in Persons Living with HIV with HIV Associated Non-AIDS Conditions. AIDS and Behavior. 2017. DOI 10.1007/s10461-017-1786-6.

[6] Dodson S, Klassen KM, McDonald K, Millard T, Osborne RH, Battersby MW, Fairley CK, Simpson JA, Lorgelly P, Tonkin A, Roney J. HealthMap: a cluster randomised trial of interactive health plans and self-management support to prevent coronary heart disease in people with HIV. BMC infectious diseases. 2016;16(1):114.

[7] Ghayomzadeh M, SeyedAlinaghi S, Shamsi MM, Rezaei S, Earnest CP, Akbarnejad S, Taj L, Mohraz M, Navalta JW, Ghasemi P, Voltarelli FA. Effect Of 8-Week of Hospital-Based Resistance Training Program on TCD4+ Cell Count and Anthropometric Characteristic of HIV Patients in Tehran, Iran: A Randomized Controlled Trial. Journal of strength and conditioning research. 2017 Dec.

[8] Yadav S. Perceived social support, hope, and quality of life of persons living with HIV/AIDS: a case study from Nepal. Qual Life Res. 2010;19(2):157–66.

[9] Negin J, Nyirenda M, Seeley J, Mutevedzi P. Inequality in health status among older adults in Africa: The surprising impact of anti-retroviral treatment. Journal of Cross-Cultural Gerontology 2013 28: 491–493. PMID: 24122525

[10] Nyirenda M, Chatterji S, Falkingham J, Mutevedzi P, Hosegood V, Evandrou M, et al. An investigation of factors associated with the health and well-being of HIV-infected or HIV-affected older people in rural South Africa. BMC Public Health 2012 12: 259. doi: 10.1186/1471-2458-12-259 PMID: 22471743

[11] Kuteesa MO, Seeley J, Cumming RG, Negin J. Older people living with HIV in Uganda: understanding their experience and needs. African Journal of AIDS Research 2012 11(4): 295–305. doi: 10.2989/16085906.2012.754829 PMID: 25860188

[12] Millard T, Agius PA, McDonald K, Slavin S, Girdler S, Elliott JH. The positive outlook study: a randomised controlled trial evaluating online self-management for HIV positive gay men. AIDS and Behavior. 2016;20(9):1907-18.

[13] Feldman MB, Arakaki LS, Raker AR. Participation in a Self-management Intervention for People Living With HIV. Journal of the Association of Nurses in AIDS Care. 2016 Jul 1;27(4):530-7.

[14] Russell S, Martin F, Zalwango F, Namukwaya S, Nalugya R, Muhumuza R, Katongole J, Seeley J. Finding meaning: HIV self-management and wellbeing among people taking antiretroviral therapy in Uganda. PloS one. 2016;11(1):e0147896.

[15] Chu C, Selwyn PA. An epidemic in evolution: the need for new models of HIV care in the chronic disease era. Journal of Urban Health. 2011;88(3):556-66.

[16] Webel AR, Moore SM, Hanson JE, Salata RA. The rationale, design, and initial efficacy of SystemCHANGE™-HIV: A systems-based intervention to improve physical activity in people living with HIV. Journal of AIDS & clinical research. 2013;4(3).

[17] Modeste RR, Majeke SJ. Self-care symptom-management strategies Amongst women living with HIV/AIDS in an urban area in KwaZulu-Natal. Health SA Gesondheid. 2010;15(1):1-8.

[18] Wang K, Chen WT, Zhang L, Bao M, Zhao H, Lu H. Facilitators of and barriers to HIV self-management: Perspectives of HIV-positive women in China. Applied Nursing Research. 2016 Nov 30;32:91-7.
[19] Niakan S, Mehraeen E, Noori T, Gozali E. Web and Mobile Based HIV Prevention and Intervention Programs Pros and Cons – A Review. Studies in Health Technology and Informmatics. 2017; 236:319-327.
[20] Wagner N, Ouedraogo D, Artavia-Mora L, Bedi A, Thiombiano BA. Protocol for a randomized controlled trial evaluating mobile text messaging to promote retention and adherence to antiretroviral therapy for people living with HIV in Burkina Faso. JMIR research protocols. 2016;5(3):e170) doi:10.2196/resprot.5823
[21] Brown J, Hanson JE, Schmotzer B, Webel AR. Spirituality and optimism: a holistic approach to component-based, self-management treatment for HIV. Journal of religion and health. 2014;53(5):1317-28.
[22] Johnson MO, Dilworth SE, Taylor JM, Neilands TB. Improving coping skills for self-management of treatment side effects can reduce antiretroviral medication nonadherence among people living with HIV. Annals of Behavioral Medicine. 2011;41(1):83-91.
[23] Wang K, Chen WT, Zhang L, Bao M, Zhao H, Lu H. Facilitators of and barriers to HIV self-management: Perspectives of HIV-positive women in China. Applied Nursing Research. 2016;32:91-7.
[24] Fan V, Duran D, Silverman R, Glassman A. HIV/AIDs intervention packages in five countries: a review of budget data. Center for Global Development Policy Paper. 2013; (33).
[25] Nematollahi M, Khalesi N, Moghaddasi H. A Comparative Study of Management Information System for HIV / AIDS in selected countries. Journal of Payesh 2012; 11(4): 425-433.
[26] Millard T, et al. Informing the development of an online self-management program for men living with HIV: a needs assessment. BMC Public Health 2014; 14(1209): 1-9.
[27] Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: Focus group findings from HIV-positive persons and young mothers. International journal of medical informatics. 2013;82(4):e38-46.
[28] Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a Self-management Program on Patients with Chronic Disease. Effective Clinical Practice .2001; 4(6):256–62.
[29] Donovan JL, Blake DR. Qualitative study of interpretation of reassurance among patients attending rheumatology clinics: “just a touch of arthritis, doctor?” BMJ .2000; 320:541–44.
[30] Emamzadeh-Fard S, E Fard S, SeyedAlinaghi S, Paydary K. Adherence to anti-retroviral therapy and its determinants in HIV/AIDS patients: a review. Infectious Disorders-Drug Targets (Formerly Current Drug Targets-Infectious Disorders). 2012;12(5):346-56.
[31] Banerjee AK, Narasu ML. Can mobile phones transform healthcare in low and middle income countries? BMJ 2015;350:h1975.
[32] Nhavoto JA, Gronlund A, Chaquilla WP. SMSaude: Design, Development, and Implementation of a Remote/Mobile Patient Management System to Improve Retention in Care for HIV/AIDS and Tuberculosis Patients. JMIR Mhealth Uhealth 2015; 3(1): e26.
[33] Blackstock OJ, Patel VV, Cunningham CO. Use of technology for HIV prevention among adolescent and adult women in the United States. Current HIV/AIDS Reports. 2015; 12(4):489-99.
[34] Thompson MA, Mugavero MJ, Amico KR, Cargill VA, Chang LW, Gross R, Orrell C, Altice FL, Bangsberg DR, Bartlett JG, Beckwith CG. Guidelines for improving entry into and retention in care and antiretroviral adherence for persons with HIV: evidence-based recommendations from an International Association of Physicians in AIDS Care panel. Annals of internal medicine. 2012; 156(11):817-33.
[35] Tierney WM, Beck EJ, Gardner RM, Musick B, Shields M, Shiyonga NM, Spohr MH. A pragmatic approach to constructing a minimum data set for care of patients with HIV in developing countries. Journal of the American Medical Informatics Association. 2006; 13(3):253-60.
[36] Tierney WM, Rotich JK, Hannan TJ, Siika AM, Biondich PG, Mamlin BW, Nyandiko WM, Kimaiyo S, Wools-Kaloustian K, Sidle JE, Simiyu CJ. The AMPATH medical record system: creating, implementing, and sustaining an electronic medical record system to support HIV/AIDS care in western Kenya. Studies in health technology and informatics. 2007;129(1):372.
[37] McDonald K, Slavin S, Pitts MK, Elliott JH, HealthMap Project Team. Chronic disease self-management by People With HIV. Qualitative health research. 2016; 26(6):863-70.
[38] Schnall R, Bakken S, Rojas M, Travers J, Carballo-Dieguez A. mHealth technology as a persuasive tool for treatment, care and management of persons living with HIV. AIDS and Behavior. 2015;19(2):81-9.
[39] Martin S,Chinnock P, Perales J, LeeHY, LeeCY,Wu YT,Mavrodaris A. Self-management interventions for people living with HIV/AIDS. Cochrane Database of Systematic Reviews 2014, Issue 9. Art.No.: CD008731. DOI: 10.1002/14651858.CD008731.pub2.

[40] Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS and Behavior. 2016; 20(5):1049-59.
[41] Bärnighausen T, Chaiyachati K, Chimbindi N, Peoples A, Haberer J, Newell M-L. Interventions to increase antiretroviral adherence in sub-Saharan Africa: a systematic review of evaluation studies. The Lancet 2011 11(12): 942–951. doi: 10.1016/S1473-3099(11)70181-5 PMID: 22030332
[42] Mills EJ, Bakanda C, Birungi J, Chan K, Ford N, Cooper CL, et al. Life expectancy of persons receiving combination antiretroviral therapy in low income countries: a cohort analysis from Uganda. Annals of Internal Medicine 2011 155(4): 209–216. doi: 10.7326/0003-4819-155-4-201108160-00358 PMID: 21768555
[43] Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: A systematic review. PLoS Med 2007 4(10): e298. doi: 10.1371/journal.pmed.0040298 PMID: 17941716
[44] Ware NC, Idoko J, Kaaya S, Biraro IA, Wyatt MA, Agbaji O, et al. Explaining adherence success in Sub-Saharan Africa: An ethnographic study. PLOS Medicine 2009 6(1): e1000011. doi: 10.1371/journal.pmed.1000011
[45] Nhavoto JA, Grönlund Å, Klein GO. Mobile health treatment support intervention for HIV and tuberculosis in Mozambique: Perspectives of patients and healthcare workers. PloS one. 2017; 12(4):e0176051.
[46] DeSilva MB, Gifford AL, Keyi X, Li Z, Feng C, Brooks M, et al. Feasibility and Acceptability of a Real- Time Adherence Device among HIV-Positive IDU Patients in China. AIDS Res Treat. 2013; 2013:1±6.
[47] Millard T, Elliott J, Girdler S. Self-management education programs for people living with HIV/AIDS: a systematic review. AIDS patient care and STDs. 2013; 27(2):103-13.
[48] Rotheram-Borus MJ, Ingram BL, Swendeman D, Lee A. Adoption of self-management interventions for prevention and care. Primary Care: Clinics in Office Practice. 2012;39(4):649-60.
[49] Swendeman D, Ramanathan N, Baetscher L, Medich M, Scheffler A, Comulada WS, Estrin D. Smartphone self-monitoring to support self-management among people living with HIV: Perceived benefits and theory of change from a mixed-methods, randomized pilot study. Journal of acquired immune deficiency syndromes (1999). 2015; 69(0 1):S80.
[50] Ansari RM, Ansari SY, Al-Safarini S, Ayub M. Socio-Ecological Approach to Self-Management of HIV/AIDS: A Case Study of Women Patients with HIV Virus. Occupational Medicine & Health Affairs. 2013; 1(136). doi: 10.4172/2329-6879.1000136
[51] Westergaard RP, Genz A, Panico K, Surkan PJ, Keruly J, Hutton HE, Chang LW, Kirk GD. Acceptability of a mobile health intervention to enhance HIV care coordination for patients with substance use disorders. Addiction science & clinical practice. 2017; 12(1):11.
[52] Henry BL, Moore DJ. Preliminary Findings Describing Participant Experience With iSTEP, an mHealth Intervention to Increase Physical Activity and Improve Neurocognitive Function in People Living with HIV. Journal of the Association of Nurses in AIDS Care. 2016; 27(4):495-511.
[53] Sullivan KM, Inouye J. Feasibility of a Group-Based Self-Management Program for Ethnically Diverse People With HIV/AIDS in Hawaii. Asian / Pacific Island Nursing Journal, 2015; 1(2): 1-11.
[54] L'Engle KL, Green K, Succop SM, Laar A, Wambugu S. Scaled-Up Mobile Phone Intervention for HIV Care and Treatment: Protocol for a Facility Randomized Controlled Trial. JMIR research protocols. 2015;4(1):e11.
[55] Dejman M, Ardakani HM, Malekafzali B, Moradi G, Gouya MM, Shushtari ZJ, Alinaghi SA, Mohraz M. Psychological, Social, and Familial Problems of People Living with HIV/AIDS in Iran: A Qualitative Study. International Journal of Preventive Medicine. 2015;6:126. doi: 10.4103/2008-7802.172540. eCollection 2015.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-148

Telemedicine in Diagnosis, Treatment and Management of Diseases in Children

Abbas SHEIKHTAHERI[a,b], Farzaneh KERMANI[a,1]
[a] *Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran*
[b] *Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran*

Abstract. The purpose of this study was to review different telemedicine services in diagnosis, treatment and management of various children diseases and providing an overview of systematic reviews conducted in this regard. We searched English articles published in peer-reviewed journals between 2000 to 2016. We found that tele-pediatric services have been reported in various areas such as cardiology, burn, diabetes, obesity, emergency medicine, speech and hearing loss, Ear, Nose and Throat, psychology and psychiatry, radiology, oncology, home healthcare, asthma, genetics and dentistry. These studies mainly reported positive results. However, systematic reviews in tele-pediatric showed that these studies have not proven the clinical effectiveness or suggested further studies to assess the clinical outcomes of services provided through telemedicine technologies.

Keywords. telemedicine, telehealth, child, pediatrics

1. Introduction

According to American Telemedicine Association, telemedicine is "the use of medical information exchanged from one site to another via electronic communications to improve a patient's clinical health status" [1]. Tele-pediatric (application of telemedicine for children) is used in the areas of medicine in which the distance is an issue, and includes services such as diagnosis, treatment, prevention of diseases, care providers'/patients' education, research and care evaluation [2]. Pediatric services typically include care provided to children from birth to 18 years of age [3]. The purpose of this study was to review and introduce different tele-pediatric services and the consequences of using this type of services and providing an overview of systematic reviews conducted in this regard.

2. Method

A combination of keywords such as "children", "child", "pediatric", "telehealth", "telecare", "telemedicine", "tele", "review" and "systematic review" were searched in different databases including "PubMed/Medline", "Science Direct", "Web of Science" and "Cochrane library". In addition, we searched specialized journals in the field of telemedicine and considered the related references in selected articles. The inclusion

[1] Corresponding author, Farzaneh Kermani, Department of Health Information Management, Iran University of Medical Sciences, Tehran, Iran; Email: Kermani-f@tak.iums.ac.ir.

criteria were: 1) published English papers in peer-reviewed journals between 2000 to 2016, 2) papers reported a telemedicine service for children only or parents of children. In addition, the exclusion criteria were: 1) Articles focused on both children and adults, unless the findings were reported separately in the article or more than 50% of participations were children. 2) Infant and newborn community. After screening the titles and abstracts and removing the unrelated papers, full text of included papers were reviewed. The review has been done by two independent reviewers. If there were disagreements, a discussion was conducted to reach a consensus. We classified articles according to the medical domains and discussed these areas by proving some examples of studies and relevant reported outcomes. We defined telemedicine as any healthcare services provided remotely through synchronous or asynchronous communication by any modalities and technologies to children or their parents.

3. Results and discussion

Telemedicine is used in various clinical domains for children. In the following sections, each of these areas and related literature and outcomes are discussed.

Cardiology: Pediatric tele-cardiology has been applied in different age groups. It was found that tele-echocardiography increases efficiency and the quality of care, reduces unnecessary patient transport, improves sonographer skills, saves cost, reduces the length of stay in the Coronary Care Unit (CCU) and increases physicians' and patients' satisfaction [4]. Pediatricians perceived this type of services practical and cost-effective [5]. Appropriate exchange of information between hospitals, better use of equipment, improved distance learning in terms of clinicians' perspectives and rapid diagnosis, reduced costs and access to children heart tele-consultation from the perspective of the patients have also been reported [6].

Burn: In tele-burn projects for children, store and forward and videoconferencing modes are used [7-8]. According to studies, tele-burn services for children leads to cost savings of patient travel and patients' convenient access to specialized services [9]. The quality of information collected through videoconferencing was the same as the in-person visits and agreement in consultation between these two methods was reported 84% [10].

Diabetes: According to the studies, telemedicine was in accordance with the in person visits, and reduces the time and cost as well as improves rural children's access to diabetic care [11]. Telephone consultation, videoconferencing between school nurses, child and clinical team [12], messaging [13] and web portal [14] are some of telemedicine services for diabetic children that can improve clinical outcomes (reduced Hemoglobin A1c (HbA_{1c})) [12-14], reduce unnecessary calls to care centers, reduce visits in Emergency Department (ED) and hospital admissions [12].

Obesity: The use of sensors to assess physical activity and dietary habits [15], tele-consultation services [16], and tele-monitoring [17] are some of the telemedicine services in this domain. It was shown that both methods of telemedicine and telephone interventions are acceptable, feasible and effective in providing obesity treatment to children regarding outcomes (BMI, diet and quality of life) [18]. In addition, no difference was observed in parents' satisfaction with telemedicine vs. in person consultation [16] and telemedicine vs. telephone consultation [18].

Emergency medicine: A study showed that tele-consultation for children living in rural areas resulted in fewer medication errors than telephone consultation [19].

Telemedicine also resulted in a reduction in travel costs and emergency care costs [20]. Improving care quality, diagnosis and treatment, patients' and provider's satisfaction [21-22] have also been reported for tele-pediatrics in emergency services. Additionally, a study showed 98% and 92.5% agreement in treatment and diagnosis between in person and telemedicine services for emergency [23].

Speech and hearing loss: Tele-interventions have been used for deaf or hard-of-hearing children [24] in cases of children with stuttering problems [25] and training parents in the early stages of autism [26]. According to an RCT on deaf children, telehealth services increased cost savings [27]. Furthermore, effectiveness of videoconference services for students with speech problems and their improvement were similar to in person services [28]. The studies on students, families, speech therapy pathologists and school principals also showed a high satisfaction with telehealth services [29]. A review showed that the use of telehealth had a positive effect on children's speech-language but the evidence in this area is still low and not enough to influence clinical practices and policy development [30].

Ear, Nose, Throat (ENT): Agreement on synchronous [31] and asynchronous methods [32] for assessment of paediatric ENT conditions were equivalent to in person method, and the agreement rate on synchronous mode is better [31]. The same result has been reported for tele-screening of hearing problems of primary school children [33]. Some studies also suggest the effectiveness of videoconference services in assessing ENT status [34]. The reduced average waiting time of referral (73 to 29 days) and the more economical nature of videoconference-based tele-ENT consultation has been also reported [35].

Psychiatry and psychology: Diagnostic outcomes and psychology problem assessment for children via videoconference, telephone and email was effective [36]. Reduced cost in rural centers by tele-psychiatry [37] and high parents' and children's satisfaction with tele-psychiatry has been reported [38]. An RCT on the effectiveness of videoconferencing on training skills of parents of children with attention deficit hyperactivity disorder showed that the treatment and parents' education was similar to in person method [39]. Additionally, mobile services to treat anxiety [40] and interactive video services for the treatment of depression [41] indicated the success of these methods.

Radiology: The use of tele-radiology has been reported in education of physicians, students and radiologists through web-based videoconferencing [42-43], and the detection of radiological images with Smartphone [44]. In a study, diagnostic accuracy of CT and MRI images of children by a mobile was 97.52% [44]. Improved access, avoid unnecessary travels, saving costs and improving outcomes due to rapid reporting and intervention have been reported in children tele-radiology [44-45].

Oncology: Services such as psychology interventions and stress reduction for children with cancer [46-47], neuro-oncology consultation via email, requesting for visual communication, answering neuro-oncology questions [48], videoconference for the diagnosis of brain tumors [49] have been reported. Tele-support services after discharge has brought families' high satisfaction [50]. Furthermore, the agreement between diagnosis via telemedicine and the routine method was 90.6% [51]. Implementation of home care after discharge with telephone consultation had a major effect on meeting the needs of children with cancer and reduced unplanned hospitalizations [52].

Home care: These services have been used in the areas such as consultation for palliative care [53-54], web-based telemedicine to reduce hospitalizations and sudden death in children with heart problems [55], telemedicine-based robotic rehabilitation

services for children with joint damage and cerebral palsy [56]. Implementation of palliative care at home with internet and telephone consultation showed that this service was feasible and acceptable and reduced imposed responsibilities on families. Satisfaction and quality of life was similar to the control group [53]. A videoconference-based home care program for children with cancer showed parents' satisfaction and reduced the level of children concern with no effect on increased costs [57].

Asthma: Peak expiratory flows test, asthma control test, daily asthma diaries, adherence to treatment, and quality of life have been assessed [58-59]. An RCT on a web application of monitoring and education indicated improved self-management skills and children's quality of life [59]. A study on tele-monitoring at home and teaching children through a website showed that inhalation (94% vs. 89%), adherence to daily asthma diaries (35.4% vs. 20.8%) in telemedicine group was better than the control group [58].

Genetics: Using tele-genetics has been reported in the areas such as tele-consultation and remote education [60-63]. Although parents perceived tele-genetic consultations positive, they preferred face to face services [61]. Additionally, the most important challenge of tele-genetics is reimbursement and cost [63]. A review showed that in tele-genetics studies, costs have not been measured officially and most studies only have pointed out the effect of tele-genetics on saving travel cost and time [64].

Dentistry: Teledentistry can be used in online education, periodontics, oral pathology, oral medicine, orthodontics, detection of dental caries [65-66]. It's very common for dental caries treatments in childhood [67] and screening programs in schools [66]. Comparison between in-person examination and teledentistry showed positive results [68]. Also, this method is reliable versus in person screening [66, 69]. Overall, teledentistry resulted in saved time, money and travel [70], improved the care quality [65] and facilitated the timely treatment by early diagnosis [71].

Systematic reviews of telehealth in pediatrics: In many areas, no systematic review is conducted and the number of current systematic reviews is also limited. A systematic review and meta-analysis by reviewing 10 RCTs showed that telemedicine had no effect on the HbA_{1c} level, severe hypoglycemia or diabetic ketoacidosis in children with type 1 diabetes. Based on this review, limited data has been reported on patients' satisfaction, quality of life and cost. The authors concluded that telemedicine benefits may not be adequately reported [72]. Another review of 13 studies on the patient outcomes in the management of children obesity showed that the interventions are effective on increasing screening for BMI and weight management [73]. A review of the tele-pediatrics in ED showed that the positive effects of these interventions could not be supported because of little evidence [74]. A scoping review on 23 studies about the management of children with hearing problems showed that most studies have focused on hearing screening and these services improved access and coverage in rural and remote areas and less attention is paid to diagnostic interventions and rehabilitation services [29]. Another review of rehabilitation of children with hearing loss showed that the effect of the use of online technologies on promoting learning in deaf children is positive but only four studies were eligible [75].

A review of telemedicine in psychology showed that the studies have addressed feasibility, cost-effectiveness, and the patients'/providers' satisfaction and positive results were reported in all three domains. However, authors concluded that few studies have been conducted on clinical outcomes [76]. Another review on eight studies related to children with autism showed that telemedicine services were mainly implemented for behavioral and diagnostic assessment, training consultation, and monitoring behavioral interventions and telemedicine had positive effects on treating children with autism, but

more RCTs in real locations are needed [77]. Another review of 33 studies related to tele-home care showed that there were only six studies about children. This review showed that tele-home care has the ability to improve services and outcomes such as supportive roles for families, reduced stress, enhanced communication between physicians and families and reduced unplanned admissions [78]. A meta-analysis of nine studies on the effect of web or telephone-based self-care education of children with asthma suggested this intervention resulted in fewer school absences, reduced ED visits and hospitalizations [79]. In a systematic review, tele-genetics services in the field of children and adults showed that this service is suitable to provide consultation and follow-ups and has the capability of assessment and diagnosis of children suspected of genetic diseases [64]. Another review on teledentistry showed that education, diagnosis, consultation and treatment are common services in this domain [80]. Based on another review, teledentistry has an acceptable diagnostic performance in the detection of dental caries. However, more studies on effectiveness of teledentistry to caries detection are needed [81].

4. Conclusion

Studies have shown that the use of tele-pediatrics in most cases results in saving cost, reducing unnecessary travels, patients' and providers' satisfaction, and better patient self-management. Various results in terms of feasibility and acceptance, improvement of diagnosis quality and treatment outcomes, clinical effectiveness, improved accuracy of diagnosis, efficiency and quality of care have been reported. However, few systematic reviews were reported regarding tele-pediatrics, that most of them have shown some of these benefits. Most of original researches have shown positive clinical results; however, the results of systematic reviews except for limited cases have not confirmed positive clinical outcomes or suggested further studies to assess the clinical outcomes of tele-pediatric services. We aimed to introduce the application areas of telepediatrics and reported outcomes not conducting a systematic review to analyze these outcomes. Therefore, we have limitations in these regard. However, in short, we can conclude that there is little evidence about clinical outcomes and effectiveness of telemedicine services for children. Therefore, conducting further systematic reviews of the specific applications of tele-pediatrics is recommended. Also it is suggested that in order to thorough implementation and long-term usage, evaluation of telemedicine systems have been done in the real environment and with a large number of users.

References

[1] B.L. Burke, R.W. Hall, P.J. Dehnel, et al. Telemedicine: Pediatric applications, *Pediatrics* 136(1) (2015), e293-e308.
[2] R. Wootton, J. Batch, *Telepediatrics: Telemedicine and child health,* Royal Society of Medicine Press, 2005.
[3] B. McCrossan, F. Casey, *Telemedicine in the diagnosis and management of congenital heart disease*, Intech, 2011.
[4] A. Krishnan, M. Fuska, R. Dixon R, et al, The evolution of pediatric tele-echocardiography: 15-year experience of over 10,000 transmissions, *Telemed J E Health* 20(8) (2014), 681-686.
[5] B.A. McCrossan, N.N. Doherty, A.J. Sands, et al, Survey of paediatricians' opinions on a regional paediatric telecardiology service, *J Paediatr Child Health* 50(6) (2014), 482-486.

[6] L.V. Lapão, A. Correia, Improving access to pediatric cardiology in Cape Verde via a collaborative international telemedicine service, *Stud Health Technol Inform* (2015), 51-57.
[7] A.C. Smith, J.A. Kairl, R. Kimble, Post-acute care for a paediatric burns patient in regional Queensland, *J Telemed Telecare* 8(5) (2002), 302-304.
[8] A.C. Smith, K.Youngberry, J. Mill, et al, A review of three years experience using email and videoconferencing for the delivery of post-acute burns care to children in Queensland, *Burns* 30(3) (2004), 248-252.
[9] A.C. Smith, K. Youngberry, F. Christie, et al, The family costs of attending hospital outpatient appointments via videoconference and in person, *J Telemed Telecare* 9(2_suppl) (2003), S58-61.
[10] A.C. Smith, R. Kimble, J. Mill, et al. Diagnostic accuracy of and patient satisfaction with telemedicine for the follow-up of paediatric burns patients, *J Telemed Telecare* 10(4) (2004), 193-198.
[11] C.L. Wood, S.A. Clements, K. McFann, et al, Use of telemedicine to improve adherence to american diabetes association standards in pediatric type 1 diabetes, *Diabetes Technol Ther* 18(1) (2016), 7-14.
[12] R. Izquierdo, P.C. Morin, K. Bratt, et al, School-centered telemedicine for children with type 1 diabetes mellitus, *J Pediatr* 155(3) (2009), 374-379.
[13] B. Bin-Abbas, M. Jabbari, A. Al-Fares, et al, Effect of mobile phone short text messages on glycaemic control in children with type 1 diabetes, *J Telemed Telecare* 20(3) (2014), 153-156.
[14] J.E. Pinsker, C. Nguyen, S. Young S, et al, A pilot project for improving paediatric diabetes outcomes using a website: The Pediatric Diabetes Education Portal, *J Telemed Telecare* 17(5) (2011), 226-230.
[15] R. Schiel, A. Kaps, G. Bieber, et al, Identification of determinants for weight reduction in overweight and obese children and adolescents, *J Telemed Telecare* 16(7) (2010), 368-373.
[16] K.W. Mulgrew, U. Shaikh, J. Nettiksimmons, Comparison of parent satisfaction with care for childhood obesity delivered face-to-face and by telemedicine, *Telemed J E Health* 17(5) (2011), 383-387.
[17] M.B. Irby, K.A. Boles, C. Jordan, et al, TeleFIT: Adapting a multidisciplinary, tertiary-care pediatric obesity clinic to rural populations, *Telemed J E Health* 18(3) (2012), 247-249.
[18] A.M. Davis, M. Sampilo, K.S. Gallagher, et al, Treating rural paediatric obesity through telemedicine vs. telephone: Outcomes from a cluster randomized controlled trial, *J Telemed Telecare* 22(2) (2016), 86-95.
[19] M. Dharmar, N. Kuppermann, P.S. Romano, et al, Telemedicine consultations and medication errors in rural emergency departments, *Pediatrics* 132(6) (2013), 1090-1097.
[20] S. Desai, M.L. Williams, A.C. Smith, Teleconsultation from a secondary hospital for paediatric emergencies occurring at rural hospitals in Queensland, *J Telemed Telecare* 19(7) (2013), 405-410.
[21] M. Dharmar, P.S. Romano, N. Kuppermann, et al, Impact of critical care telemedicine consultations on children in rural emergency departments, *Crit Care Med* 41 (2013), 2388-2395.
[22] B. Heath, R. Salerno, A. Hopkins, et al, Pediatric critical care telemedicine in rural underserved emergency departments, *Pediatr Crit Care Med* 10(5) (2009), 588-591.
[23] M. Raska, Mathison. D, telemedicine for acute care: Compairing in-person visits to telemedicine visits at a pediatric emergency department, *Int J Telerehabil* 7(2) (2015).
[24] M. McCarthy, K. Muñoz, K.R. White, Teleintervention for infants and young children who are deaf or hard-of-hearing, *Pediatrics* 126(Supplement 1) (2010), 52-58.
[25] C. Sicotte, P. Lehoux, J. Fortier-Blanc, et al, Feasibility and outcome evaluation of a telemedicine application in speech-language pathology, *J Telemed Telecare* 9(5) (2003), 253-258.
[26] E. Baharav, C. Reiser, Using telepractice in parent training in early autism, *Telemed J E Health* 16(6) (2010), 727-731.
[27] K.M. Blaiser, D. Behl, C. Callow-Heusser, et al. Measuring costs and outcomes of tele-intervention when serving families of children who are deaf/hard-of-hearing, *Int J Telerehabil* 5(2) (2013), 3-10.
[28] S. Grogan-Johnson, R. Alvares, L. Rowan, et al A pilot study comparing the effectiveness of speech language therapy provided by telemedicine with conventional on-site therapy, *J Telemed Telecare* 16(3) (2010),134-39.
[29] S.M. Govender, M. Mars, The use of telehealth services to facilitate audiological management for children: A scoping review and content analysis, *J Telemed Telecare* 23(3) (2016), 392-401.
[30] O.D. Taylor, N.R. Armfield, P. Dodrill, et al. A review of the efficacy and effectiveness of using telehealth for paediatric speech and language assessment, *J Telemed Telecare* 20(7) (2014), 405-412.
[31] A.C. Smith, S. Dowthwaite, J. Agnew, et al, Concordance between real-time telemedicine assessments and face-to-face consultations in paediatric otolaryngology, *Med J Aust* 188(8) (2008), 457-460.
[32] A.C. Smith, C. Perry, J. Agnew, et al, Accuracy of pre-recorded video images for the assessment of rural indigenous children with ear, nose and throat conditions, *J Telemed Telecare* 12(3_suppl) (2006), 76-80.
[33] P. Lancaster, M. Krumm, J. Ribera, et al, Remote hearing screenings via telehealth in a rural elementary school, *Am J Audiol* 17(2) (2008), 114-122.

[34] A.C. Smith, N.R. Armfield, W.I. Wu, et al, A mobile telemedicine-enabled ear screening service for Indigenous children in Queensland: Activity and outcomes in the first three years, *J Telemed Telecare* 18(8) (2012), 485-489.
[35] C.Q. Xu, A.C. Smith, P.A. Scuffham, et al, A cost minimisation analysis of a telepaediatric otolaryngology service, *BMC Health Serv Res* 8(1) (2008).
[36] P.M. Yellowlees, D.M. Hilty, S.L. Marks, et al, A retrospective analysis of a child and adolescent eMental Health program, *J Am Acad Child Adolesc Psychiatry* 47(1) (2008), 103-107.
[37] R. Spaulding, N. Belz, S. DeLurgio, et al, Cost savings of telemedicine utilization for child psychiatry in a rural Kansas community, *Telemed J E Health* 16(8) (2010), 867-871.
[38] K.M. Myers, J.M. Valentine, S.M. Melzer, Child and adolescent telepsychiatry: Utilization and satisfaction, *Telemed J E Health* 14(2) (2008),131-137.
[39] Y. Xie, J.F. Dixon, O.M. Yee, et al, A study on the effectiveness of videoconferencing on teaching parent training skills to parents of children with ADHD, *Telemed J E Health* 19(3) (2013), 192-199.
[40] G. Pramana, B. Parmanto, P.C. Kendall, et al, The SmartCAT: An m-health platform for ecological momentary intervention in child anxiety treatment, *Telemed J E Health* 20(5) (2014), 419-427.
[41] E.L. Nelson, M. Barnard, S. Cain, Feasibility of telemedicine intervention for childhood depression, *Counselling and psychotherapy research* 6(3) (2006), 191-195.
[42] A.M. Monteiro, D.G. Corrêa, A.A. Santos, et al, Telemedicine and pediatric radiology: A new environment for training, learning, and interactive discussions, *Telemed J E Health* 17(10) (2011), 753-756.
[43] A.B. Silva, A.C. de Amorim, A Brazilian educational experiment: Teleradiology on web TV, *J Telemed Telecare* 15 (2009), 373-376.
[44] F. Zennaro, M. Bava, A. Casalino, et al, Smartphones for mobile teleradiology in paediatric imaging: Evaluation of diagnostic accuracy, *Egyptian Computer Science Journal* 37(5) (2013), 94-98.
[45] R, Griggs, S, Andronikou, R. Nell, et al, World federation of pediatric imaging [WFPI] volunteer outreach through tele-reading: the pilot project in South Africa, *Pediatr Radiol* 44(6) (2014), 648-654.
[46] S.D. Hutchison, H. Sargeant, B.A. Morris, et al, A community-based approach to cancer counselling for patients and carers: A preliminary study, *Psychooncology* 20(8) (2011), 897-901.
[47] D.C. Seitz, C. Knaevelsrud, G. Duran, et al, Efficacy of an internet-based cognitive-behavioral intervention for long-term survivors of pediatric cancer: A pilot study, *Support Care Cancer* 22(8) (2014), 2075-2083.
[48] I. Qaddoumi, E. Bouffet, Supplementation of a successful pediatric neuro-oncology telemedicine-based twinning program by e-mails, *Telemed J E Health* 15(10) (2009), 975-982.
[49] I. Qaddoumi, A. Mansour, A. Musharbash, et al, Impact of telemedicine on pediatric neuro-oncology in a developing country: The Jordanian-Canadian experience, *Pediatr Blood Cancer* 48(1) (2007), 39-43.
[50] M. Bensink, N. Armfield, H. Irving, et al, A pilot study of videotelephone-based support for newly diagnosed paediatric oncology patients and their families, *J Telemed Telecare* 14(6) (2008), 315-321.
[51] T.C. Santiago, J.J. Jenkins, F. Pedrosa, et al, Improving the histopathologic diagnosis of pediatric malignancies in a low-resource setting by combining focused training and telepathology strategies, *Pediatr Blood Cancer* 59(2) (2012), 221-225.
[52] M.C. Yilmaz, S.A. Ozsoy, Effectiveness of a discharge-planning program and home visits for meeting the physical care needs of children with cancer, *Support Care Cancer* 18(2) (2010), 243-253.
[53] N. Bradford, J. Young, N.R. Armfield, et al, A pilot study of the effectiveness of home teleconsultations in paediatric palliative care, *J Telemed Telecare* 18(8) (2012), 438-442.
[54] N.K. Bradford, N.R. Armfield NR, J. Young, et al, Principles of a paediatric palliative care consultation can be achieved with home telemedicine, *J Telemed Telecare* 20(7) (2014), 360-364.
[55] A.K. Black, U.K. Sadanala, C.E. Mascio, et al, Challenges in implementing a pediatric cardiovascular home telehealth project, *Telemed J E Health* 20(9) (2014), 858-867.
[56] K. Chen, Y. Ren, D. Gaebler-Spira, et al, Home-based tele-assisted robotic rehabilitation of joint impairments in children with cerebral palsy, Conf Proc IEEE, 2014, 5288-5291.
[57] H. Hansson, H. Kjaergaard, C. Johansen, et al, Hospital-based home care for children with cancer: Feasibility and psychosocial impact on children and their families,*Pediatr Blood Cancer*60(5)(2013),865-72.
[58] D.S. Chan, C.W. Callahan, V.B. Hatch-Pigott, et al, Internet-based home monitoring and education of children with asthma is comparable to ideal office-based care: Results of a 1-year asthma in-home monitoring trial, *Pediatrics* 119(3) (2007), 569-578.
[59] R.L. Jan, J.Y. Wang, M.C. Huang, et al, An internet-based interactive telemonitoring system for improving childhood asthma outcomes in Taiwan, *Telemed J E Health* 13(3) (2007), 257-268.
[60] D. Gustafson, M. Wise, A. Bhattacharya, et al, The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: A randomized controlled trial, *J Med Internet Res* 14 (2012), e101.

[61] B. Hopper, M. Buckman, M. Edwards, Evaluation of satisfaction of parents with the use of videoconferencing for a pediatric genetic consultation, *Twin Res Hum Genet* 14(4) (2011), 343-346.
[62] L.P. Maximino, M.M. Picolini-Pereira, J.L. Carvalho, Telegenetics: application of a tele-education program in genetic syndromes for Brazilian students, *J Appl Oral Sci* 22(6) (2014), 477-483.
[63] H.J. Stalker, R. Wilson, H. McCune, et al, Telegenetic medicine: Improved access to services in an underserved area, *J Telemed Telecare* 12(4) (2006), 182-185.
[64] J.S. Hilgart, J.A. Hayward, B. Coles, et al, Telegenetics: A systematic review of telemedicine in genetics services, *Genet Med* 14(9) (2012), 765-776.
[65] A.Bhambal, S. Saxena, S.V. Balsaraf, Teledentistry: potentials unexplored,*J Int Oral Health* 2(3) (2010), 1-6.
[66] S, Bissessur, A comparative analysis of traditional dental screening versus tele dentistry screening. 2016.
[67] D.T. Kopycka-Kedzierawski, R. Billings, Prevalence of dental caries and dental care utilisation in preschool urban children enrolled in a comparative-effectiveness study, *Eur Arch Paediatr Dent* 12(3) (2011), 133-138.
[68] K. Adlassnig, Remote diagnosis of children dental problems based on non-invasive photographs–a valid proceeding?, *Stud Health Technol Inform*, 150 (2009), 458-462.
[69] T. Subbalekshmi, V. Anandan, R. Apathsakayan, Use of a Teledentistry-based Program for Screening of Early Childhood Caries in a School Setting, *Cureus* 9(7) (2017), 1-7.
[70] K.C. Pentapati, P. Mishra, M. Damania, et al, Reliability of intra-oral camera using teledentistry in screening of oral diseases–Pilot study, *Saudi Dent J* 29(2) (2017), 74-77.
[71] D.T. Kopycka-Kedzierawski, R.J. Billings, Teledentistry in inner-city child-care centres, *J telemed telecare* 12 (4) (2006), 176-181.
[72] R.M. Shulman, C.S. O'Gorman, M.R. Palmert, The impact of telemedicine interventions involving routine transmission of blood glucose data with clinician feedback on metabolic control in youth with type 1 diabetes: A systematic review and meta-analysis, *Int J Pediatr endocrinol* 2010(1) (2010), 1-9.
[73] A,J, Smith, Á. Skow, J. Bodurtha, et al, Health information technology in screening and treatment of child obesity: A systematic review, *Pediatrics* 131(3) (2013), 894-902.
[74] R. Gattu, G. Teshome, R. Lichenstein, Telemedicine applications for the pediatric emergency medicine: A Review of the current literature, *Pediatr Emerg Care* 32(2) (2016), 123-130.
[75] A. Simpson, A. El-Refaie, C. Stephenson, et al, Computer-based rehabilitation for developing speech and language in hearing-impaired children: A systematic review, *Deafness Educ Int* 17(2) (2015), 111-119.
[76] J. Van Allen, A.M. Davis, S. Lassen, The use of telemedicine in pediatric psychology: Research review and current applications, *Child Adolesc Psychiatr Clin N Am* 20(1) (2011), 55-66.
[77] M. Boisvert, R. Lang, M. Andrianopoulos, et al, Telepractice in the assessment and treatment of individuals with autism spectrum disorders: A systematic review, *Dev Neurorehabil* 13(6) (2010), 423-432.
[78] N, Bradford, N.R. Armfield, J. Young, et al, The case for home based telehealth in pediatric palliative care: A systematic review, *BMC Palliat Care* 12(1) (2013), 1-13.
[79] E. Ahmad, D.E. Grimes, The effects of self-management education for school-age children on asthma morbidity: a systematic review, *J Sch Nurs* 27(4) (2011), 282-292.
[80] R. Mariño, A. Ghanim, Teledentistry:A systematic review of the literature, *J Telemed Telecare* 19(4) (2013), 179-183
[81] M. Estai, S. Bunt, Y. Kanagasingam, et al, Diagnostic accuracy of teledentistry in the detection of dental caries: A systematic review, J *Evid Based Dent Pract* 16(3) (2016), 161-172.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-156

Towards an IMU Evaluation Framework for Human Body Tracking

Verena VENEK[a,1], Wolfgang KREMSER[a] and Cornelia SCHNEIDER[a]
[a]*Salzburg Research Forschungsgesellschaft mbH, Salzburg, Austria*

Abstract. Existing full-body tracking systems, which use Inertial Measurement Units (IMUs) as sensing unit, require expert knowledge for setup and data collection. Thus, the daily application for human body tracking is difficult. In particular, in the field of active and assisted living (AAL), tracking human movements would enable novel insights not only into the quantity but also into the quality of human movement, for example by monitoring functional training. While the current market offers a wide range of products with vastly different properties, literature lacks guidelines for choosing IMUs for body tracking applications. Therefore, this paper introduces developments towards an IMU evaluation framework for human body tracking which compares IMUs against five requirement areas that consider device features and data quality. The data quality is assessed by conducting a static and a dynamic error analysis. In a first application to four IMUs of different component consumption, the IMU evaluation framework convinced as promising tool for IMU selection.

Keywords. IMU, human body tracking, evaluation framework.

1. Introduction

In recent years, sensor miniaturization has been a driver for product innovation in many areas. The miniaturization of Inertial Measurement Units (IMUs) for instance enables numerous applications including the tracking of moving objects by computing orientation [1,2]. In the field of human movement analysis, the integration of IMUs in wearable full-body tracking systems has been realized by a few specialized manufacturers and start-ups [3], such as Xsens (Xsens Technologies B.V., Enschede, The Netherlands), Enflux (Enflux, San Francisco, United States) or Rokoko (Rokoko Electronics, Copenhagen, Denmark). In general, all these systems aim at assessing people's movement outside the laboratory [4,5]. In particular, in the research field of Active and Assisted Living (AAL), mobility for older people means independence and participation in social life. Furthermore, the functional ability of being mobile, including exercise, promotes healthy aging and postpones frailty [6]. In order to avoid injury and maintain functional ability, it is important that exercises are performed in a proper way. Exercise monitoring of older people at their homes could benefit from wearing an IMU-based full-body tracking system. The quality assessment of exercise performance could not only provide meaningful feedback to the users themselves but it could also help supervising entities like trainers or doctors to adjust training according to the users' abilities. Functional training has been supported technologically in various AAL projects for better quality of life and health of older people [7,8]; however, only few have considered IMUs as sensor technology so far [9,10]. The required expert knowledge of

[1] Corresponding Author: Verena Venek, Salzburg Research Forschungsgesellschaft mbH, Jakob Haringer Straße 5/3, 5020 Salzburg, E-Mail: verena.venek@salzburgresearch.at.

existing full-body tracking systems for setup and data acquisition as well as their fixed configuration of IMU placements is unsuitable for the target group of older people. Thus, we decided on implementing a human body tracking system where number, position, and hardware can be chosen according to application requirements.

Available IMUs on the market differ in various aspects including component composition. The range of products starts at basic packages, for which casing, battery and data transmission have to be added separately, up to turnkey solutions. Some of the turnkey solutions are even capable of outputting their orientation directly. Due to the variety of options, IMUs have to be selected based on criteria.

Existing works stated some criteria related to IMU selection such as power consumption and sampling rate [1,11-13]. However, none considered data quality of the IMU measurements, which is a crucial factor since all further processing steps depend on it.

Thus, this paper introduces an evaluation framework for wearable IMUs used in human body tracking that considers both, hardware features and data quality. In order to assess its performance, the framework is applied to IMUs with different component composition. Based on the results of the application, possible improvements and adaptations of the framework should be identified and discussed.

2. Methods

2.1. IMU evaluation requirements

In recent years, considerably little effort has been investigated in the comparison of IMU performance with respect to human motion analysis. Ahmad et al. [1] identified form factor, data accuracy, response rate, and the degree of freedom (DOF) when selecting IMUs for various applications. Data accuracy was described to be dependent on the selected sensor fusion which reduced sensor drift and other errors introduced by the sensors. During the development of their own IMU for the analysis of Parkinson's disease symptoms, Rodríguez-Martín et al. [11] stated several IMU requirements, including long, unsupervised runtime, minimum power consumption, wearable form factor and connectivity to other devices. Within their work, they compared twelve commercial IMUs including the Xsens MTw and the Physilog 3 as wireless body tracking sensors. The conducted comparison relied purely on market analysis without looking at data samples of the actual devices. A latest design methodology for motion capture wearables, called *Octopus*, is exclusively based on review of publication and a market research [13]. It considers connection, attachment method, and the physical properties of the device such as shape, dimensions, weight, housing material and color.

Based on this research, we created a demand profile for IMUs in human body tracking. In addition, we considered a role model for the framework since the usage of such full-body tracking systems require expert knowledge, which complicates the independent usage, e.g. for AAL applications, and the predefined number and positions of IMUs decrease the flexibility of applications. Hence, the full-body tracking system that uses wireless IMUs, namely the Xsens MVN Awinda [14] is considered as role model. Finally, the identified requirements for human body tracking are divided into five areas:

Form factor: Wearable IMUs for wireless body tracking are available in different dimensions. Nevertheless, they should be appropriate for unobtrusive integration into

clothes. The Xsens MVN Awinda uses Velcro straps to attach the IMUs onto predefined body segments. Each IMU comes with dimensions of *47x30x13mm* and weights *16g* [15]. Integration can be additionally possible using clips or other simple methods of fastening [1,13].

Mobility: The maximum possible operating time depends on battery capacity, sampling rate and the means of data transmission. A runtime of *6 h* at *60 Hz* with wireless transmission is considered as threshold for the framework since high-end systems like Xsens MVN Awinda work with these specifications [14]. Ideally, the battery is rechargeable via Universal Serial Bus (USB).

Data acquisition: For immediate feedback to users, near real-time data processing should be targeted, i.e. showing a latency time of at most *100ms* [16]. Furthermore, wireless data transmission should be considered for avoiding cables and, thus, difficult setup. For example, Bluetooth Low Energy (BLE) is most widely supported by consumer smartphones and meets the requirement of fast data transmission. The supported sampling rate of the IMU should be at least *60 Hz* , being higher than the suggestion of *40 Hz* by [11]. An API for monitoring the hardware's health status and for customization of applications is a basic requirement. This should enable access to the IMU's raw data.

Additional features: The price of each IMU should be of concern since we plan to develop a full-body tracking system to consist of at least five IMUs [4]. Considering the current prices of smart wearables such as smartphones, we expect that people might be willing to spend the same amount for a wearable IMU-based full-body tracking system, being priced up to €500 [17], i.e. the price of a single IMU should not exceed *€ 100*. In addition, the possibility of building a sensor network should be possible so that multiple IMUs can interact with each other for synchronization purposes.

Reliable recognition of human motion: This requirement area includes the verification of how robust and reproducible the data provided by the IMU is for full-body tracking. Most commonly, IMUs with 6-DOF, which comprise a 3D accelerometer and a 3D gyroscope, or 9-DOF IMUs, which add a 3D magnetometer, are used [1]. Thus, at least 6-DOF IMUs should be required. Further sensors, like the magnetometer or the barometer are optional for consideration in orientation estimation. With respect to data quality, a static and dynamic error analysis support the evaluation of IMUs by identifying error behavior of the IMUs and rank the devices upon these results (see subsection *2.2*).

The more requirements of the framework are fulfilled by the IMU under investigation, the more appropriate it should be for the task of body tracking.

2.2. Static and dynamic error analysis

Static and dynamic error analysis were used for the objective evaluation of the data quality of each IMU.

The **static error analysis** aims at evaluating the amount of inherent noise in the sensor readings, particularly, the gyroscope drift error. For a static and undisturbed position, the IMU is fixed onto a table and data is recorded for *120 min* without movement at room temperature. If available, the IMU's internal calibration routine is used to dampen the noise. To quantify the residual noise occurring while the sensor experiences no movement, an Allan variance analysis is conducted as described in El-Sheimy et al. [18] determining two types of error quantity: the random walk and the bias stability. The Angle Random Walk (ARW) or Velocity Random Walk (VRW) is the

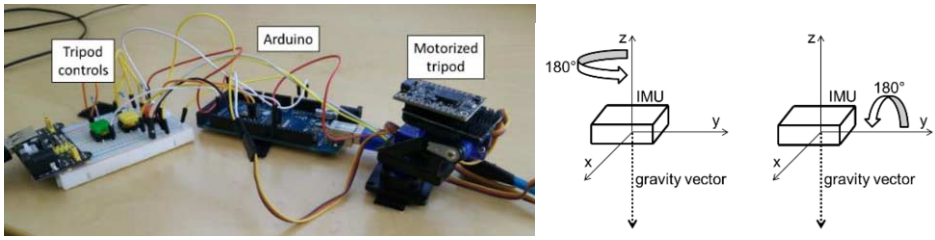


Figure 1. Left: Motorized tripod (on which the device under test is mounted) with additional Arduino micro controller unit and manual controls; Right: IMU rotations about z- and y-axis.

influence of high-frequency noise. The bias stability is dependent on the influence of low-frequency noise mainly coming from the sensor's electronics. With the bias changing over time, this metric tells the best expected bias stability (BS) from this IMU. Of both quantities, the root mean square error (RMS) over x-, y- and z-axis is calculated, for which lower numbers imply better results.

By conducting a **dynamic error analysis**, the performance of the sensors in motion is assessed and compared to simulated data. To move the IMU in a controlled manner, a motorized pan-tilt tripod and an Arduino board is used to conduct counterclockwise z- and y-rotations of *180°* (see *Figure 1*). All rotations are performed at a programmed speed of *166.7°/s* that roughly resembles a natural limb rotation. To mark beginning and end of the sequence, rotations at maximum speed of the motors around the z-axis are executed giving clearly distinguishable impulses in the data. If necessary, the IMU's frames are mapped to a right-handed coordinate system, where the positive vertical z-axis points upwards, the positive y-axis to the right, and the positive x-axis frontwards. Sensor fusion algorithm such as described by [19] are used to calculate quaternions representing the orientation of each IMU over time. For comparison, the entire motion sequence is simulated using the spherical linear interpolation (SLERP) of quaternions which is commonly used to smoothly animate 3D rotations [20]. The calculated and simulated quaternions are applied to rotate a unit vector, resulting in an ideal and an actual trajectory. The idea is to determine the error by using the distance function $\phi(q_1, q_2)$ that computes the angular deviation between two quaternions [21]:

$$\phi(q_1, q_2) = \cos^{-1}(2\langle q_1,\ q_2\rangle^2 - 1) \tag{1}$$

The denotation $\langle q_1,\ q_2\rangle$ corresponds to the inner product of the two quaternions q_1 and q_2. The range of the distance function is between *0* and π, where *0* means that the compared rotations result in the same orientation. Values close to π represent maximal angular deviation from simulated orientation.

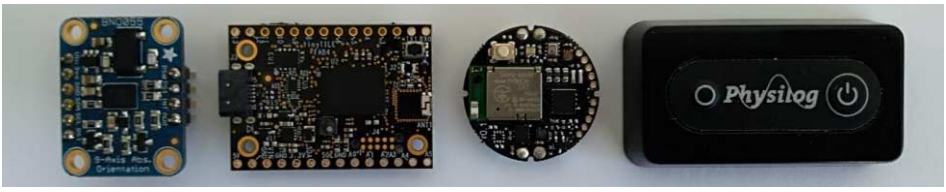


Figure 2. IMUs of different components composition from left to right: BNO055 board, Intel Curie board, MetaMotionC and Physilog 5.

3. Results

In this section the first application of the framework is presented showing the fulfilments of the requirements and the ranking of the static and dynamic error analysis (*Table 1*). Thus, four IMUs were selected based on their component composition. The BNO055 board [22] and the Intel Curie board [23] represented basic packages that can be integrated into low-level prototypes. In this case, "low-level" means that the IMU requires additional hardware, like power supply and casing. The MetaMotionC [24] and Physilog 5 [25] represented turnkey solutions that work out of the box and come with complementary software. We chose these devices due to their small size, wide availability, and their differing component composition.

Form factor: The dimensions of all four devices are sufficient (*Figure 2*). However, the BNO055 and Curie provide no attachment to the body. This could be circumvented by using 3D printed parts. For the MetaMotionC, the 3D-printed case provided by the manufacturer was used. Physilog provides a rubber clip with each device.

Mobility: The Physilog comes with a USB-chargeable battery included in the casing. BNO055 and Curie have to be powered externally, although the Curie provides a charging circuit. The MetaMotionC does not provide an integrated rechargeable battery and is instead powered by a coin cell.

Data acquisition: Except BNO055, which supports only *I²C* communication, all investigated IMUs support BLE transmission. Regarding API, libraries are freely available for all boards except Physilog.

Additional features: The Physilog is the most expensive one of the selected IMUs with *€ 499* per device, which would sum up to *€ 2,495* for a full-body tracking system. All others were within the required price range. The support for multiple sensors is limited. Solely Physilog and MetaMotionC provide basic synchronization by allowing IMUs to start and stop simultaneously.

Table 1. Application of IMU evaluation framework to four IMUs: 'X' marks fulfilment of requirement and the rank of each IMU is given from *1* (best) to *4* for static and dynamic error analysis.

	Physilog 5	MetaMotionC	BNO055	Intel Curie
Form Factor				
Application to clothes	X	X		
Size < *47x30x13 mm*	X	X	X	X
Weight < *16 g*	X	X	X	X
Mobility				
Battery life > *6 h*	X	X		
Chargeable battery	X	X		
USB charging	X			X
Data acquisition				
Bluetooth data transmission	X	X		X
Sampling rate > *60 Hz*	X	X	X	X
API available		X	X	X
Additional features				
Price less than *€ 100*		X	X	X
Multiple sensor network feasibility	X	X		
Reliable recognition				
6-DOF IMU	X	X	X	X
3-DOF magnetometer	X	X	X	
Barometer	X	X		
Static error analysis – RW	*2*	*1*	*4*	*3*
Static error analysis – BS	*1*	*2*	*3*	*4*
Dynamic error analysis	*1*	*2*	*3*	*4*

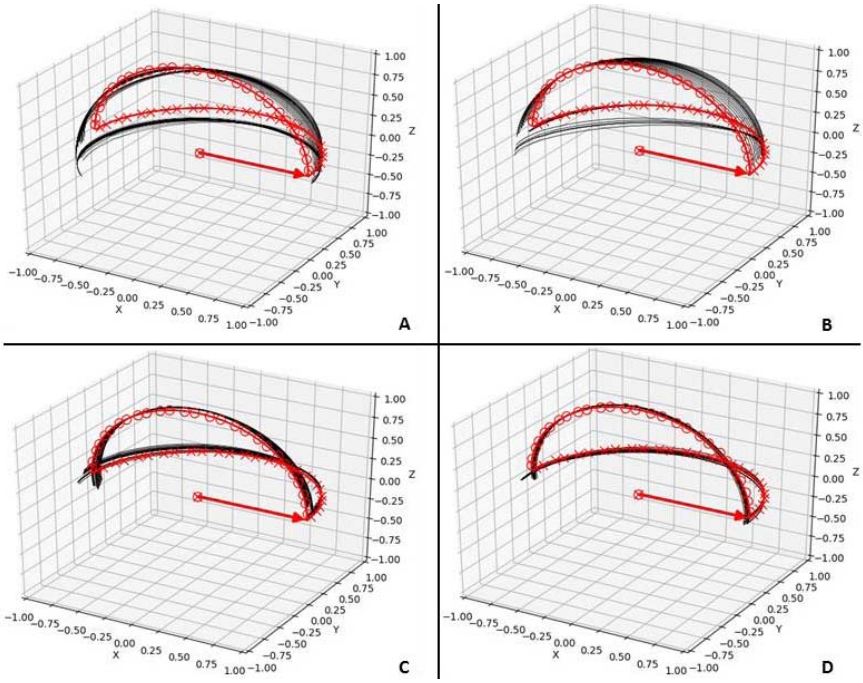


Figure 3. IMU-based trajectory of unit vector of (A) BNO055, (B) Intel Curie, (C) MetaMotionC and (D) Physilog 5 (black) and SLERP-simulated trajectory of the same unit vector (red marked with X and O).

Reliable recognition of human motion: All four selected IMUs provide the mandatory 3-DOF gyroscope and accelerometer. Additional sensors are available on the Physilog, MetaMotionC and BNO055. The results of the static and dynamic error analysis are given in *Table 2*. The best results related to bias stability came from the Physilog measurements (RMS for gyroscope and accelerometer of *7.19 °/h* and *0.12 °/h*, respectively). MetaMotionC provided the lowest amount of random walk (RMS for ARW and VRW of *0.0055 °/s/√Hz* and *0.0001°/s/√Hz*, respectively). For the dynamic error analysis, the calculated orientation was compared to the orientation data simulated with SLERP (*Figure 3*). The resulting maximal and mean angular orientation deviations are shown in *Table 2*. The lowest deviation was achieved by using the Physilog data, with the MetaMotionC coming close. The BNO055 was ranked third, and the largest error by a clear margin came from fusing the Curie data (see *Table 1)*.

Table 2. *Static error analysis*: RMS of random walk in °/s/√Hz and of bias stability in °/h over the x-, y- and z-axis; *Dynamic error analysis*: maximum and mean angular orientation derivation to SLERP-simulated orientation data ranging from *0* to π (*= 3.14*).

	Physilog 5	MetaMotionC	BNO055	Intel Curie
Bias stability				
Gyroscope:	7.19	7.43	7.95	7.54
Accelerometer:	0.12	0.27	0.29	0.50
Random walk				
ARW:	0.0068	0.0055	0.0455	0.0069
VRW:	0.0002	0.0001	0.0002	0.0002
$\Phi(q_1, q_2)$				
Maximum:	1.12	1.19	1.58	3.14
Mean:	0.35	0.37	0.43	1.09

4. Discussion and conclusions

The introduced IMU evaluation framework was applied to four IMUs with different component composition. Results indicated that the turnkey solutions MetaMotionC and Physilog are better due to device features and data quality. Considering the basic packages BNO055 and Curie, which ranked behind the turnkey solutions, they differed extremely from the other two products in their component composition, i.e. battery and casing would have to be added additionally. However, the framework could even distinguish between the basic packages and the turnkey solutions by identifying the missing components when observing the requirements in the areas mobility and data transmission.

The binary marking of fulfilments gave an overview of how often each IMU meets the requirements. The two measures random walk error and bias stability as part of the static error analysis of the framework for evaluating data quality proved to be useful and in accordance with the results of the dynamic error analysis. With respect to the dynamic error analysis, the novel usage of SLERP does not require special equipment and provides at the same time reliable reference data, although, it is important to mention that the generated reference orientation is idealized.

If required for a more detailed evaluation, weighting of the requirements could easily be included. This would allow emphasis on application-specific aspects, for example, favoring mobility over form factor. In addition, the framework could be extended by defining exclusion criteria for IMUs as a first step, such as minimal component composition. If required, application-specific thresholds for the error quantities of the static error analysis could be added as requirements. Furthermore, in a framework extension, it would make sense to rate bias stability more important than random walk since it relates to the non-linear gyroscope drift, while random walk can be reduced by applying additional filters in the orientation estimation process.

The current IMU evaluation framework provides useful requirements and assessment methods to evaluate IMUs objectively for human body tracking. Particularly, the orientation simulation by using SLERP proved suitable for a first assessment of the IMU's performance during dynamic motions.

The next step will be the development of a full-body tracking system based on results of the IMU evaluation framework. This system will be used for the application in AAL projects for monitoring physical activity. Particularly, in order to provide an advanced support for older people to maintain their functional abilities as long as possible, the quality of exercise at home will be monitored considering, for example, back and leg axis stability.

References

[1] N. Ahmad, R. Ghazilla, N. Khairi, and V. Kasi, Reviews on Various Inertial Measurement Unit (IMU) Sensor Applications, Int. J. Signal Proccesing Syst.1(2) (2013), 256–262.

[2] M. B. del Rosario, S. J. Redmond, and N. H. Lovell, Tracking the evolution of smartphone sensing for monitoring human movement, Sensors (Switzerland) 15(8) (2015), 18901–18933.

[3] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, Survey of Motion Tracking Methods Based on Inertial Sensors: A Focus on Upper Limb Human Motion, Sensors 17(6) (2017), 1257.

[4] F. J. Wouda, M. Giuberti, G. Bellusci, and P. H. Veltink, Estimation of Full-Body Poses Using Only Five Inertial Sensors : An Eager or Lazy Learning Approach ?, Sensors (Basel) 16(12) (2016), 2138.

[5] D. Roetenberg, H. Luinge, and P. Slycke, Xsens MVN : Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors, Tech. Rep.(2013), 1–7.
[6] H. Mollenkopf, The significance of out-of-home mobility in modern society, in: Enhancing Mobility in Later Life, H. Mollenkopf, F. Marcellini, I. Ruoppila, Z. Széman, and M. Tacken, Eds. IOS Press, 2005, chapter 1.
[7] GameUp, http://www.gameupproject.eu/, last access: 29-Jan-2018.
[8] ZentrAAL, https://www.zentraal.at/, last access: 29-Jan-2018.
[9] Active@Home, http://www.active-at-home.com/, last access: 29-Jan-2018.
[10] Co-Train, https://www.cotrain.eu/, last access: 29-Jan-2018.
[11] D. Rodríguez-Martín, C. Pérez-López, A. Samà, J. Cabestany, and A. Català, A wearable inertial measurement unit for long-term monitoring in the dependency care area, Sensors (Basel) 13(10) (2013), 14079–14104.
[12] A. Ebert, M. Kiermeier, C. Marouane, and C. Linnhoff-Popien, SensX: About Sensing and Assessment of Complex Human Motion, Proceedings of 14th IEEE International Conference on Networking, Sensing and Control (ICNSC) (2017).
[13] J. Marin, T. Blanco, and J. J. Marin, Octopus: A design methodology for motion capture wearables, Sensors (Switzerland) 17(8) (2017), 1–24.
[14] Xsens, Xsens MVN Awinda, https://www.xsens.com/products/xsens-mvn-analyze, last access: 17-Jan-2018.
[15] Xsens, Xsens MTw Awinda, https://www.xsens.com/products/mtw-awinda, last access: 17-Jan-2018.
[16] J. Nielsen, Usability engineering. Elsevier, 1994.
[17] Statista, Umfrage zum angemessenen Preis für ein neues Smartphone in Deutschland, https://de.statista.com/statistik/daten/studie/722561/umfrage/umfrage-zum-angemessenen-preis-fuer-ein-neues-smartphone-in-deutschland/, last access: 24-Jan-2018.
[18] N. El-Sheimy, H. Hou, and X. Niu, Analysis and Modeling of Inertial Sensors Using Allan Variance, IEEE Trans. Instrum. Meas. 57(1) (2008), 140–149.
[19] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, Estimation of IMU and MARG orientation using a gradient descent algorithm, in 2011 IEEE International Conference on Rehabilitation Robotics, (2011), 1–7.
[20] E. B. Dam, M. Koch, and M. Lillholm, Quaternions , Interpolation and Animation, Tech. Rep. DIKU-TR-98/5 (1998), 103.
[21] D. Q. Huynh, Metrics for 3D rotations: Comparison and analysis, J. Math. Imaging Vis. 35(2) (2009), 155–164.
[22] Bosch Sensortec, BNO055, https://www.bosch-sensortec.com/bst/products/bno055, last access:18-Jan-2018.
[23] L. Premier Farnell, tinyTILE - Intel Curie Module Based Board, https://www.element14.com/community/docs/DOC-82913/l/tinytile-intel-curie-module-based-board, last access: 27-Jun-2017.
[24] mbientlab, MetaMotionC, https://mbientlab.com/product/metamotionc/, last access: 27-Jun-2017.
[25] GaitUp, Physilog(R)5: motion sensor, https://shop.gaitup.com/hardware/1-physilog5.html, last access: 27-Jun-2017.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-164

Automated Error Detection in Physiotherapy Training

Marko JOVANOVIĆ[a,1], Johannes SEIFFARTH[a], Ekaterina KUTAFINA[a,b] and Stephan M. JONAS[a]
[a]*Department of Medical Informatics, Uniklinik RWTH Aachen, Aachen, Germany*
[b]*Faculty of Applied Mathematics, AGH University of Science and Technology, Cracow, Poland*

Abstract. *Background*: Manual skills teaching, such as physiotherapy education, requires immediate teacher feedback for the students during the learning process, which to date can only be performed by expert trainers. *Objectives*: A machine-learning system trained only on correct performances to classify and score performed movements, to identify sources of errors in the movement and give feedback to the learner. *Methods*: We acquire IMU and sEMG sensor data from a commercial-grade wearable device and construct an HMM-based model for gesture classification, scoring and feedback giving. We evaluate the model on publicly available and self-generated data of an exemplary movement pattern executions. *Results*: The model achieves an overall accuracy of 90.71% on the public dataset and 98.9% on our dataset. An AUC of 0.99 for the ROC of the scoring method could be achieved to discriminate between correct and untrained incorrect executions. *Conclusion*: The proposed system demonstrated its suitability for scoring and feedback in manual skills training.

Keywords. Wearable Technology, Gestures, Machine Learning, Feedback, Education, mHealth.

1. Introduction

Teaching of practical and motoric manual skills in the education of physiotherapy students is traditionally done in a classroom setting through observation and repetition. It is essential for the physiotherapy students to receive immediate teacher feedback during the learning process. This allows for early correction of mistakes and optimization of students' competences. Challenges are limited face to face time with teachers as well as shortage of personnel. This motivates the need for computer-supported systems to monitor and assess students' performance.

In general, methods for gesture recognition can be divided into computer vision-based methods and wearable-based methods, such as sensing gloves or forearm armbands. Computer vision-based methods usually require a specific setup and are limited in mobility. Thus, we are focusing on wearable-based methods here. Body-worn sensors typically include inertial measurement units (IMU), surface electromyography (sEMG) sensors and haptic sensors. In the past, sEMG has been used in conjunction with different machine learning algorithms for applications in sign language recognition [1], human computer interaction [2], virtual training [3], and prosthetics [4]. However, only limited work has been done in the usage of wearable sensors in training of manual tasks.

[1] Corresponding Author: Marko Jovanović, Department of Medical Informatics, Uniklinik RWTH Aachen, Pauwelsstr. 30, 52074 Aachen, Germany, E-Mail: mjovanovic@mi.rwth-aachen.de

Kutafina et al. utilize a combination of Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) in a wearable-based eLearning system for the support of manual skills training in hand hygiene [3]. Forearm armbands equipped with an IMU and sEMG sensors detect the user's hand motions and evaluate the performance on hand hygiene tasks, achieving a recognition rate of 98.30%.

Here, we present a novel wearable-based method for automated assessment and evaluation of physiotherapy performance in a teaching setting. The wearable sensor armband Myo by Thalmic Inc. is used to monitor the performance of physiotherapy students in training. The method is explicitly not created for evaluation of a patient's therapy adherence. The Myo is a Bluetooth-connected, consumer-grade human activity recognition (HAR) device, containing an 8-channel sEMG sensor (200 Hz sampling rate) and a 9-axis IMU (50 Hz sampling rate). A recent validation of EMG data collected with the Myo armband has shown that it is comparable to more costly EMG setups [5]. The data acquired by the sensor armband is analyzed using a Hidden Markov Model (HMM), similarly to approaches used in speech recognition [6]. Based on the HMM performance, a feedback in the form of a performance score is given.

While our method is being trained solely on correct data, it is still able to detect incorrect samples and describe errors in performed gestures. This contrasts with standard classifier training, where each error type needs to be trained as a separate class, which requires extensive training data and knowledge about all possible errors.

2. Methods

To investigate the feasibility of giving automated feedback on the performance of manual skills training in physiotherapy education, following steps had to be performed: (i) acquisition and preprocessing of training datasets; (ii) development and training of a movement model; (iii) development of error detection in the movement model, trained only on positive data; and (iv) evaluation of the method.

2.1. Used Datasets and Feature Extraction

For method training and evaluation, we acquire data from one subject performing the *Flexion-Abduction-External Rotation with Elbow Flexion* (FAEREF) pattern, as an exemplary proprioceptive neuromuscular facilitation (PNF) pattern. PNF is a common treatment concept in physiotherapy.

In a total of 7 recording sessions, we acquired raw sensor data from all sensors of two Myo devices worn both arms of a subject. Four sessions numbered $C1$ to $C4$ contain data of correctly performed movements (with $N = 11, 11, 11,$ and 14 repetitions respectively), while three sessions contain data of erroneous executions for each errors typically seen during training of the FAEREF pattern: ($E1$) the patient's arm is moved besides its body instead of over and behind its head ($N = 13$), ($E2$) the therapist does not follow the movement by a rotation of his upper body ($N = 12$), and ($E3$) the arm of the patient is not moved diagonally and the initial position of the patient's hand is wrong ($N = 15$).

Secondly, in order to have the method evaluated on a larger and more general dataset we utilize a publicly available dataset of basketball referee gestures by Yeh et al. [7]. The dataset contains accelerometer and EMG data from one Myo device for 14 different

basketball referee gestures ($G1$ to $G9$ arm gestures, $G11$ to $G15$ finger gestures) performed 10 times by 11 subjects.

For construction of the feature vectors used in training and evaluation, IMU acceleration and gyroscope raw data are used without further preprocessing. The EMG data provided by the Myo armband reflects forearm muscle tissue activity and highly depends on subject muscle tissue, skin moisture, fat and body hair. As preprocessing, (i) full-wave rectification [7], (ii) downsampling to 50 Hz, and (iii) synchronization with IMU data is performed. Finally, the data is normalized across all 8 EMG channels, such that only relative muscle activity is used.

2.2. Training/Construction of the Movement Model

We use the datasets for training of an HMM-based movement model, used both as a movement classifier and for error identification in the movement trained only on correct physiotherapy performances.

2.2.1. Movement Modeling with HMMs

The exemplary FAEREF pattern can be divided into six independent submovements (Figure 1). Therefore, we start by constructing an HMM [6], introducing one state for each submovement ($S = \{1, \dots, 6\}$). We impose following restrictions on the system: (i) submovements can only start after completion of the previous submovement, (ii) the order of execution is fixed, and, (iii) a movement always starts with the first and ends with the last submovement. Thus, the HMM allows only transitions to the next state and repetitions within one state. The transition probabilities, are determined by the duration of each submovement (Figure 1).

Submovements are considered to be hidden in the sense that during movement execution only sensor readings and derived features can be acquired, which allow to uncover the "hidden" underlying submovement states. We further specify the HMM by training it with observation sequences recorded solely from movement data of correctly performed executions, separately for the left and the right arm, and obtained from physiotherapy professionals. We iteratively perform segmental K-means algorithm [8] to optimize the HMM to the most appropriate description of the training set of observation sequences and denote such model a movement HMM (mHMM). Since performing physiotherapy typically involves both the left and right arm of the performer, we further introduce a derived type of the trained HMM: the two-arms movement HMM (2-mHMM) $\lambda = (\lambda_1, \dots, \lambda_X) = \left(\left(\lambda_{1,L}, \lambda_{1,R}\right), \dots, \left(\lambda_{X,L}, \lambda_{X,R}\right)\right)$, that models the 2-arm movements $1, \dots, X$ where $\lambda_{X,L}$ and $\lambda_{X,R}$ denote the mHMMs for the left and right arm of movement X, respectively.

2.3. Movement Evaluation

Based on the measured observation sequence as well as the trained model, the state sequence that maximizes the probability of the observations is used as the basis for computing the most likely segmentation of the performed movement execution.

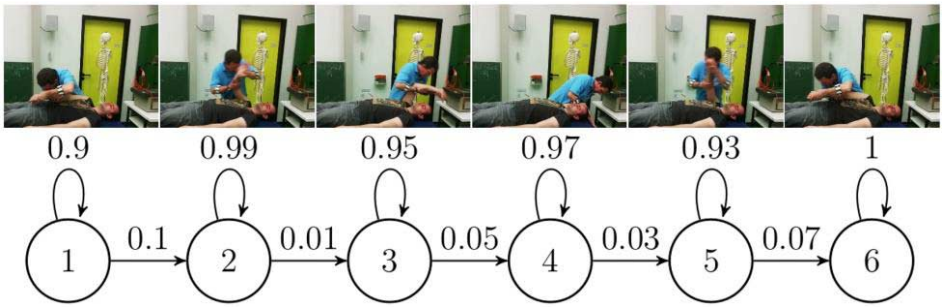


Figure 1. Upper part: Performance of the FAEREF movement pattern, consisting of six individual submovements. While performing the pattern, the physiotherapist moves the arm of the patient from a hip position diagonally over and behind the patient's head. During the execution, the upper body of the therapist rotates and follows her or his arms. Lower part: The HMM states and transition probabilities modelling the submovements of the FAREF movement pattern.

2.3.1. Classification of Movements

Given an observation of a single arm movement and a set of trained mHMMs for left and right arm movement, the most likely path through the individual submovements is calculated using the Viterbi path finding algorithm [9]. For a pair of arm movement observation sequences (left and right arm), and given trained 2-mHMMs, the classification is based on the most likely path probabilities.

The result of this step is a classification of the 50 Hz feature vector into one class of the six sub-movements per observation. However, this is not sufficient for generating feedback on the quality of the performed movement. Specifically, the model did only learn correct movements, therefore, all classes are "correct" sub-movements and sources of error cannot be identified based on class information. Therefore, an addition to the classification process is made.

2.3.2. Defining Tolerance Regions for Performance Acceptance

As the movement executions can be erroneous only within a portion of the whole arm movement, we focus on identifying abnormality in the observations in comparison to the expected class behaviour. Thus, we define tolerance regions for the observations of a feature. Feature data within the tolerance region will be accepted as being emitted from a correct execution, whereas feature data outside the tolerance region is considered to be emitted from erroneous executions. For multivariate Gaussian distributions, the tolerance regions are hyperellipsoids that contain $100\beta\%$ of the distribution population with a significance level of $100(1-\alpha)\%$. The tolerance hyperellipsoid for a p-variate Gaussian distributed data is given by the estimates of the mean mu and covariance matrix Sigma and the size l of the sample data. The tolerance factor $c \in R$ is found such that all x that fulfill the condition that the Mahalanobis distance [10] of all such x is smaller than c:

$$(x-\mu)^T \Sigma^{-1} (x-\mu) \leq c \tag{1}$$

are contained within the tolerance region. A suitable approximation of the tolerance factor is given by the harmonic mean [11]:

$$c = \frac{g\chi^2\left(\beta;p,\frac{p}{n}\right)}{(\chi^2(\alpha;e))} \tag{2}$$

where $g = g(p,n)$, $e = e(p,n)$ and $\chi^2(\beta;p,p/n)$ is the β -th percentile of a non-central χ^2 distribution with k degrees of freedom and the non-centrality parameter p/n. We compute tolerance factors $c_{i,\phi}$ for each submovement i and feature ϕ, allowing to distinguish between feature data from correct and erroneous executions. The features acceleration, orientation and gyroscope have 3-variate Gaussian distributions, whereas the EMG feature has an 8-variate distribution. For computation, we have used the parameters $\beta = 0.95$ and $\alpha = 0.05$. Further, we define the tolerance region

$$Tol_{i,\phi} = \left\{ o_\phi \middle| \left(o_\phi - \mu_{i,\phi}\right)\Sigma_{i,\phi}^{-1}\left(o_\phi - \mu_{i,\phi}\right)^T \leq c_{i,\phi} \right\} \tag{3}$$

where o_ϕ denotes the corresponding feature data from the observation vector o.
The feature feat in the observation o_t is assumed to originate from a correct execution if the observed feature data lies within the tolerance region of the corresponding submovement:

$$(o_t)_\phi \in Tol_{i,\phi} \text{ and } o_t \in S_i(\boldsymbol{q}^*) \tag{4}$$

where $\boldsymbol{q}^*$ is the optimum state sequence. If Eq. (4) does not hold true, the feature data is assumed to originate from an erroneous execution and is further denoted as an error event. Using appropriate visualizations of the error events, the affected features can be identified over time, allowing for feedback resolved over feature and time.

2.3.3. Performance Scoring

For scoring of the movement performance, we count the number of occurrences of error events for all features and sampling points. Given F independent features (EMG and IMU features), $e_1, \ldots, e_F$ and $c_1, \ldots, c_F$ count the error events and correct events in each feature respectively. We introduce the global relative frequency of error event occurrences:

$$s = 100 \cdot \left(\sum_{i=1}^{F} c_i\right) / \left(\sum_{i=1}^{F}(c_i + e_i)\right) \tag{5}$$

which indicates the correctness of the performed movement as a whole. Thus, lower scores indicate many detected error events.

2.4. Evaluation of the Proposed Method

In order to validate the proposed method in terms of classification quality and suitability for error detection in movement data, we performed the following three experiments:

- To show the general applicability of the method on hand and arm gesture detection, the proposed model is trained as a traditional multi-class classifier and evaluated on the basketball referee dataset and compared to Yeh et al. [7].
- We further evaluate the method on the FAEREF movement pattern by variating parameters and finding the optimal feature and parameter set for the classifier.
- The scoring method as an indicator for correct and erroneous movement executions is evaluated based on the receiver operating characteristic curve.

3. Results

3.1. Evaluation on Basketball Referee Gestures Dataset

We tested mHMM with number of states of $N = 5, 6, 7$ and 10. Although a larger number of mixture distributions allows for more precise modeling, the amount of feature data for estimating the mixture densities is a limiting factor due to the risk of overfitting. Thus, we decided to test $M = 1$ and $M = 3$ Gaussian densities in the mixture, representing no and moderate mixture modeling, respectively. Further, the evaluation is performed separately for the arm movements $G1$ to $G9$ and the finger movements $G11$ to $G15$ using a 5-fold cross validation. Best results have been achieved using an HMM with $M = 3$ mixture distributions, and $N = 10$ states trained in 5 iterations (Table 1).

3.2. Evaluation of the FAEREF Movement Pattern

We evaluated the performance of a trained 2-mHMM using a 2-fold cross validation, with different model parameters, variating the state number and feature sets used for training. We trained models for each correct $C1$ to $C4$ and erroneous movements $E1$ to $E3$ and evaluated on each set. A reasonable expectation is to find confusion in the classification for correct executions (e.g., $C1$ labeled as $C2$). However, classification accuracy with the full feature set on the set of classes $C1$ to $C4$ is about 95%, which suggests that the 2-mHMM suffers from overfitting (Table 2). We further evaluated the accuracy of a common 2-mHMM using a 2-fold cross validation for all correct movement recordings $C1$ to $C4$, effectively doubling the number of training data for training of the correct models. After testing, the orientation feature appears to be the main reason for a high recognition rate, an indicator for overfitting (Table 2). Further results will therefore be reported with and without the orientation feature.

3.3. Scoring Method Evaluation

We evaluated the performance of a trained 2-mHMM using a 2-fold cross validation, with different model parameters, variating the state number and feature sets used for training. The introduced scoring method can only be considered helpful to the student if it is able to distinguish correct from erroneous movement executions. Thus, we treat the score as a binary classifier and evaluate the influence of training data and selection of feature sets on the ability of binary classification using the scoring approach of the mHMM. All combinations of two correct execution sets from $C1$ to $C4$ of the FAEREF pattern are used for training 2-mHMMs with $N = 10$ states and $M = 1$ mixture distributions. The remaining two correct execution and three erroneous execution sets $E1$ to $E3$ were used for evaluation by score on the corresponding 2-mHMM.

Table 1. Classification accuracy the method on the dataset of basketball referee gestures

Subset	Full Feature Set	Acceleration	EMG	Yeh et al. [7]
Finger movements (G1-G9)	83.84%	90.71%	76.77%	97.9%
Finger gestures (G11-G15)	43.8%	43.1%	63.1%	N/A

Table 2. Classification accuracy of 2-mHMM on the FAEREF pattern

Classifier	Full Set	Acceleration	Orientation	Gyroscope	EMG
Individual 2-mHMM	90%	70%	95%	70%	57%
Common 2-mHMM	98.9%	97.7%	100%	100%	74.7%

Based on the sensitivity and specificity of correct and erroneous executions, a receiver operating characteristic (ROC) curve (Figure 2) was created for both the (1) full feature set and (2) feature set without the orientation feature. The resulting area under curve (AUC) is 0.66, and 0.99 for the feature set with and without orientation.

4. Discussion and Outlook

Although the proposed method has not been explicitly developed with classification in mind, an evaluation based on the basketball referee gesture dataset allows to compare the performance with the existing movement classification method of Yeh et al. on the same dataset. Our best result (90.7%) comes close to their SVM-based approach which achieved an accuracy of 97.9% [7]. The achieved accuracies on the dataset suggest that the model can be generalized to a broader set of gestures and is able to distinguish between different movement types.

Further evaluation of the model on the exemplary FAEREF movement pattern indicated overfitting. The 2-fold cross validation performed for all single features reveals the orientation feature as the single responsible one for the observed overfitting, suggesting that the orientation data is not well modeled by the states. We attribute this to the geometrical background of the feature. Usage of von Mises distribution for modeling the feature data can deal with this problem [12].

Without the orientation feature, the proposed scoring method was able to discriminate highly effectively between correct and incorrect data, while being trained only on correct data (AUC of .99). This is also the major advantage of the proposed method: training relies solemnly on correct data while the model is still being able to reliably detect erroneous performances through scoring. In addition, since the scoring can identify features responsible for a low result, the source of an error can be automatically detected and reported on (e.g., rotation false or at a wrong point in time).

However, the proposed method was yet only tested with one test subject performing the PNF technique. Future research is therefore needed to investigate whether the well-performing separation property of the scoring method holds inter-subjects training.

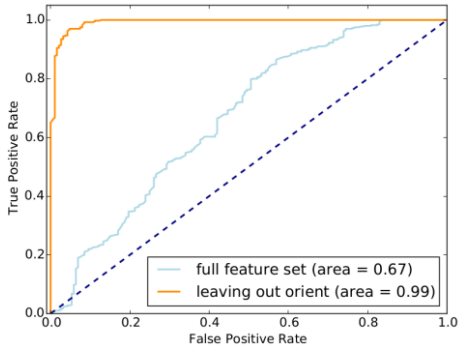


Figure 2. ROC curves for the binary classification capabilities using the feedback scoring for two different feature sets.

In addition, several aspects of the feedback system will have to be developed in the future. Providing feedback based on time-feature graphs is difficult. For example, the acceleration feature data is relative to the orientation of the Myo in the world coordinate system and the orientation feature uses the initial rotation as reference. Thus, calibration of the raw sensor data might be required.

However, the generality of the method allows for fast and simple integration of new movement models. The method is able to identify performance errors over time allowing detailed feedback generation on the performed movement and can be further adapted for usage in online evaluation in the future. The method has shown promising results for teaching PNF patterns in the training of physiotherapy and could be used in a wide range of eLearning contexts teaching manual skills.

Acknowledgments

This work was partially funded by the German Federal Ministry of Education and Research, and the European Social Fund as part of the digital media in vocational education (DIMEBB2) initiative (funding number 01P15013) and by the Faculty of Applied Mathematics AGH UST statutory tasks within subsidy of Polish Ministry of Science and Higher Education.

References

[1] Zhang X, Chen X, Li Y, Lantz V, Wang K, Yang J. A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 2011 Nov;41(6):1064–76.

[2] Kim J, Cho D, Lee KJ, Lee B. A Real-Time Pinch-to-Zoom Motion Detection by Means of a Surface EMG-Based Human-Computer Interface. Sensors (Basel). 2014 Dec 29;15(1):394–407.

[3] Kutafina E, Laukamp D, Bettermann R, Schroeder U, Jonas SM. Wearable Sensors for eLearning of Manual Tasks: Using Forearm EMG in Hand Hygiene Training. Sensors. 2016 Aug 3;16(8):1221.A.N. Author, Website Title, URL, last access: 1.1.2014.

[4] Atzori M, Müller H, Baechler M. Recognition of hand movements in a trans-radial amputated subject by sEMG. In: 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR). 2013. p. 1–5.

[5] Pizzolato S, Tagliapietra L, Cognolato M, Reggiani M, Müller H, Atzori M. Comparison of six electromyography acquisition setups on hand movement classification tasks. PLOS ONE. 2017 Oct 12;12(10):e0186132.

[6] Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE. 1989 Feb;77(2):257–86.

[7] Yeh C-W, Pan T-Y, Hu M-C. A Sensor-based Official Basketball Referee Signals Recognition System Using Deep Belief Networks. In: MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017, Proceedings, Part I [Internet]. Springer International Publishing; 2017. p. 565–575.

[8] Juang B-H, Rabiner LR. The segmental K-means algorithm for estimating parameters of hidden Markov models. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1990;38(9):1639–1641.

[9] Rabiner LR, Juang B-H. Fundamentals of speech recognition. 1993;

[10] De Maesschalck R, Jouan-Rimbaud D, Massart DL. The Mahalanobis distance. Chemometrics and intelligent laboratory systems. 2000;50(1):1–18.

[11] Krishnamoorthy K, Mathew T. Comparison of Approximation Methods for Computing Tolerance Factors for a Multivariate Normal Population. Technometrics. 1999 Aug;41(3):234–49.

[12] Mardia KV, Hughes G, Taylor CC, Singh H. A Multivariate Von Mises Distribution with Applications to Bioinformatics. The Canadian Journal of Statistics / La Revue Canadienne de Statistique. 2008;36(1):99–109.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-172

Mobile-Based Applications and Functionalities for Self-Management of People Living with HIV

Esmaeil MEHRAEEN[a], Reza SAFDARI[b,1], Niloofar MOHAMMADZADEH[b], SeyedAhmad SEYEDALINAGHI[c], Siavash FOROOTAN[a] and Minoo MOHRAZ[c]
[a]*School of Allied Medical Sciences, Tehran University of Medical Sciences, Iran*
[b]*Department of Health Information Management, Tehran University of Medical Sciences, Iran*
[c]*Iranian Research Center for HIV/AIDS, Iranian Institute for Reduction of High Risk Behaviors, Tehran University of Medical Sciences, Iran*

Abstract. Background: Due to the chronicity of HIV/AIDS and the increased number of people living with HIV (PLWH), these people need the innovative and practical approaches to take advantage of high-quality healthcare services. The objectives of this scoping review were to identify the mobile-based applications and functionalities for self-management of people living with HIV. Methods: We conducted a comprehensive search of PubMed, Scopus, Science direct, Web of Science and Embase databases for literature published from 2010 to 2017. Screening, data abstraction, and methodological quality assessment were done in duplicate. Results: Our search identified 10 common mobile-based applications and 8 functionalities of these applications for self-management of people living with HIV. According to the findings, "text-messaging" and "reminder" applications were more addressed in reviewed articles. Moreover, the results indicated that "medication adherence" was the common functionality of mobile-based applications for PLWH. Conclusion: Inclusive evidence supports the use of text messaging as a mobile-based functionality to improve medication adherence and motivational messaging. Future mobile-based applications in the healthcare industry should address additional practices such as online chatting, social conversations, physical activity intervention, and supply chain management.

Keywords. Mobile, Application, Self-management, HIV/AIDS

1. Introduction

HIV (Human Immune-deficiency Virus) is now largely considered as a chronic condition that its control is overtaking of human evolution in the healthcare industry [1]. This chronic condition is the most important reason for morbidity and mortality all over the world [2]. More than 30 million people live with HIV infection and 10 million are at Acquired Immunodeficiency Syndrome (AIDS) phase who must receive the lifesaving antiretroviral therapy (ART) [3]. Consequently, there are more people living with HIV (PLWH) (over 30 million) many of whom are qualified for ART. Nevertheless, treatment coverage is insufficient, and among the PLWH receiving antiretroviral therapy, adherence to medication is low [4]. The main issue for PLWH is providing self-

[1] Corresponding author: Reza SAFDARI, Tehran University of Medical Sciences, Qhods Square, Post Code: 7134845794, Tehran, Iran. E-mail: rsafdari@tums.ac.ir

management interventions in an active manner during their routine practices and between social and clinical visits [5,6].

In recent decades, the virtual integration of mobile technologies into routine practices has been generating many new opportunities to improve engagement in self-management interventions [7]. Self-management by mobile technologies offers new opportunities to PLWH through common functions such as appointment and medication reminders [8,9] and educational and motivational messaging [10]. The potential of mobile-based applications to provide healthcare services [11] and enhance the self-management interventions have been well documented over the recent years [12-15]. These documents cite the new mobile-based applications in the field of healthcare and management of chronic conditions [16-21]. Ramanathan et al., showed that mobile-based applications that provide telecommunication and self-management services of PLWH with a care/consultation center will provide social and psychological support for these persons and improve the quality of healthcare for them [22]. In 2016 a similar study by Garofalo et al., stated that the use of mobile applications for self-management of PLWH can ensure treatment tracking including prevention/interventions programs, medication, and clinical checkups [23].

Because of the importance of using mobile-based self-management for HIV care, more studies have been done on applications, functionalities, and challenges of this area [24-26]. With the development of this innovation, large numbers of mobile applications with a wide range of health applications have entered the healthcare market. However, use of these applications for HIV/AIDS self-management was not well-documented [27-29]. In this study, we undertook a scoping review of exploring the use of mobile-based applications and functionalities to support self-management of PLWH.

2. Methods

This study was a scoping review to identifying the mobile-based applications and functionalities for self-management of people living with HIV. We aimed to a detailed identification of different mobile-based applications and functionalities that help PLWH to manage their chronic condition. Research questions were included: 1. What are the mobile-based applications for self-management of PLWH? 2. What are the functionalities of mobile-based applications for self-management of PLWH?

2.1. Research strategy

Due to the appearance and increase of mobile-based applications to manage chronic conditions in last decade, we searched relevant databases for literature that published in 2010 to 2017. The searches were conducted in the databases PubMed, Scopus, Science direct, Web of Science and Embase. To identify published, original research and reviews that reported the mobile based applications and functionalities for self-management of HIV care, an organized search was conducted with the following search keywords in combination: Self-management, Self-care, Self-monitoring, Application, Functionality, Mobile health, m-Health, Mobile phone, Cell phone, Smartphone, and HIV/AIDS.

2.2. Inclusion & Exclusion Criteria

Our inclusion criteria were: Full-text papers with the keywords in the title or abstracts, studies that published in 2010 to November 29, 2017, and studies published in English. We excluded resources such as reports, editorial letters, newspapers, and abstracts. We also excluded studies that addressed the broader field of mobile applications, which is not applicable to the field of HIV/AIDS condition management.

3. Results

Using the applied strategies, 192 references were retrieved (for research question one 110, and research question two 82) and 29 papers were thoroughly surveyed. Table 1 shows the final analyzed articles and the studies that were found but did not meet inclusion criteria and were excluded from this paper.

Table 1. Search result from different databases

Research Question	Total Reference Retrieved	Total Duplicate References	Total Excluded References	Final Analyzed Articles
RQ1	110	32	68	10
RQ2	82	21	42	19
Total	192	53	110	29

3.1. Information extraction

In this study, to answer the research questions, reviewed articles were classified into two categories, articles that had referred to the mobile-based applications for self-management of PLWH and papers that had surveyed the functionalities of these applications. Review of first category of articles showed that there are 10 mobile-based applications for HIV self-management. Table 2 shows these applications and their functionalities with final surveyed articles. Based on the findings of this study, "Text-messaging" had the highest frequencies (6). Moreover, we obtained a set of functionalities of the mobile-based applications for self-management of PLWH. Results of this section showed that, "Medication adherence" among other functionalities, had the highest frequencies.

4. Discussion

In recent years, application of HIV/AIDS mobile-based programs in healthcare organizations and clinics has been extensively considered [30]. There is well-documented evidence of potential efficacy of mobile-based applications for addressing common challenges for people living with HIV to support communication with health care providers, increasing the ability to access services, management of mental health, reduction of substance use, a decrease in sexual risk behaviors, and enhanced medication adherence [23,24,31-33]. This scoping review aimed to detail identification of mobile-based applications and functionalities for self-management of PLWH.

Table 2. Summary of final studied articles and overview of identified apps and Functionalities

Mobile-Based Self-Management Applications											
ID	**First Author (Reference/Year)**	Health system Focused applications	AIDSinfo HIV/AIDS	Reminder applications	Hiv & Aids Guide	Facing AIDS	Patient-care focused applications	Safe sex Guide	Text-messaging	TxText tool	Aidsmap news
1	Garofalo R (23/2016)			√					√		
2	Henry BL (28/2016)			√					√		
3	Schnall R (16/2015)		√			√					
4	Forrest JI (36/2015)	√					√				
5	Ingersoll K (37/2014)									√	
6	Smillie K (31/2014)								√		
7	Odeny TA (29/2014)			√					√		
8	Muessig KE (27/2013)				√			√			√
9	da Costa TM (38/2012)			√					√		
10	Cornelius JB (34/2011)								√		
Frequency		1	1	4	1	1	1	1	6	1	1

Functionalities of Mobile-Based Self-Management Apps									
ID	**First Author (Reference/Year)**	Antiretroviral Therapy	Medication adherence	Educational messaging	attendance at appointments	Safe-sex negotiation	Facilitate communication	Reminders	Motivational messaging
11	Sharpe JD (44/2017)			√				√	
12	Cooper V (54/2017)		√		√				
13	Swendeman D (24/2016)								√
14	Nhavoto JA (19/2015)			√					√
15	Thomas B (42/2015)					√			
16	Montoya JL (48/2015)		√						√
17	Mbuagbaw L (4/2015)		√		√				
18	Tufts KA (18/2015)	√	√					√	
19	L'Engle KL (46/2015)		√					√	√
20	Shet A (45/2014)		√				√		
21	Montoya JA (50/2014)		√					√	
22	Ramanathan N (22/2013)							√	√
23	Catalani C (39/2013)			√			√	√	
24	Miller CW (49/2013)		√				√		√
25	Muessig KE (27/2013)					√	√		
26	Kalichman SC (47/2011)		√						
27	Chang LW (41/2011)			√			√		
28	Shet A (43/2010)							√	
29	Lester RT (40/2010)	√							
Frequency		2	9	4	2	2	5	7	6

We identified 10 mobile-based applications for self-management of PLWH. The use of mobile-based applications for self-management of PLWH in similar articles is well

documented [34-36]. According to the findings of this paper, "text-messaging" and "reminder" applications were more addressed in reviewed articles. Mobile-based applications can send messages daily, receive responses, and make out them to deliver personalized asserting or medication reminders [37]. Short message sending (SMS) is the primary mode of delivering an m-health application. For example, SMS could exploit on women's motivation to attend the clinic for postnatal HIV care and infant testing [23,28,29]. Use of text messaging for daily routine activities reminding represents cost-effective approach adaptations to PLWH [16,27]. According to the feedback of related studies, patients who received SMS for less than 4 months, the SMS messages aided them in treatment adherence, and they would like to continue receiving SMS messages for self-management objectives in treatment process [31,38].

With regards to the second question of this research, we obtained 8 common functionalities of the mobile-based applications for self-management of PLWH. In recent years, use of the mobile-based applications as a platform for delivering disease intervention programs and health promotion is being increased [39-41]. Integration of mobile-based applications into HIV self-management holds potential, mainly in resource-limited healthcare organizations [19,24]. Mobile-based functionalities facilitate mechanisms such as clinical and medication alerts, data collection, safe-sex negotiation, direct communication with healthcare providers (HCPs), receiving motivational [4] and educational messages, and delivery of necessary information on demand about chronic condition [22,27]. However, potential functionalities of mobile-based applications from the perspective of HCPs included: 1) enhancing patient engagement, motivation, adherence, and self-management; and 2) improving provider-patient relationships and HCP care coordination [42-44].

The result of this study indicates that "medication adherence" was the common functionality of mobile-based applications for self-management of PLWH. People living with HIV should adhere to the medication regimens because the non-adherence causes more resistance to HIV and consequently, the drugs should be prescribed in advanced doses, which leads to substantial worsening of disease, death, and increased health care costs [45-47]. Technical functionalities of mobile-based applications can be very supportive by offering reminders [8,48] and engaging patient in interventions to improve adherence in routine clinical practice [18,49]. Similar studies showed that, use of mobile technology functions such as providing a timely reminder to improve medication adherence and Engaging PLWH in their treatment may be an initial step toward enhancing healthcare services of PLWH [50-54].

5. Conclusion

In this article, we determined 10 common mobile-based applications and 8 functionalities for self-management of people living with HIV. It can be concluded that among the identified applications, "Text-messaging" with medication adherence and reminding functionalities can really support PLWH in coping with their illness. It is suggested that, more relevant studies be carried out to identify future mobile-based applications for PLWH that should address additional functionalities such as online chatting, social conversations, and physical activity intervention.

6. Acknowledgments

This paper was the result of the Ph.D. dissertation on health information management at Tehran University of Medical Sciences. The authors extend thanks to health information management professionals at Tehran University of Medical Sciences and infectious diseases specialists working at the Iranian Research Center for HIV/AIDS.

7. References

[1] Hall HI, Frazier EL, Rhodes P, Holtgrave DR, Furlow-Parmley C, Tang T, Skarbinski J. Differences in human immunodeficiency virus care and treatment among subpopulations in the United States. Journal of the American Medical Association Internal Medicine. 2013; 133(14): 1337-1344.

[2] Dalal N, Greenhalgh D, Mao X. A stochastic model of AIDS and condom use. Journal of Mathematical Analysis and Applications. 2007;325(1):36-53.

[3] Joint United Nations Programme on HIV/AIDS (UNAIDS) 2010. UNAIDS Report on the global AIDS epidemic 2010. WHO Library Cataloguing-in-Publication Data [Online; 2010, accessed in: 2017]. Available in: URL: http: www.unaids.org.

[4] Mbuagbaw L, Mursleen S, Lytvyn L, Smieja M, Dolovich L, Thabane L. Mobile phone text messaging interventions for HIV and other chronic diseases: an overview of systematic reviews and framework for evidence transfer. BMC health services research. 2015;15:33.

[5] Rotheram-Borus MJ, Ingram BL, Swendeman D, et al. Adoption of self-management interventions for prevention and care. Prim Care. 2012; 39:649–660. [PubMed: 23148958]

[6] Swendeman D, Ramanathan N, Baetscher L, Medich M, Scheffler A, Comulada WS, Estrin D. Smartphone self-monitoring to support self-management among people living with HIV: Perceived benefits and theory of change from a mixed-methods, randomized pilot study. Journal of acquired immune deficiency syndromes (1999). 2015;69(0 1):S80.

[7] Webel AR, Asher A, Cuca Y, Okonsky JG, Kaihura A, Dawson RC, et al. Measuring HIV self-management in women living with HIV/AIDS: a psychometric evaluation study of the HIV Self-management Scale. J Acquir Immune Defic Syndr. 2012; 60(3):e72-e81.

[8] Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PloS One. 2014; 9:e88166. [PubMed: 24505411]

[9] Mbuagbaw L, van der Kop ML, Lester RT, et al. Mobile phone text messages for improving adherence to antiretroviral therapy (ART): an individual patient data meta-analysis of randomised trials. BMJ Open. 2013; 3:e003950.

[10] Reback CJ, Grant DL, Fletcher JB, et al. Text messaging reduces HIV risk behaviors among methamphetamine-using men who have sex with men. AIDS Behav. 2012; 16:1993–2002. [PubMed: 22610370]

[11] Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev. 2012; 3:CD009756.

[12] Mehraeen E, Ayatollahi H, Ahmadi M. Health Information Security in Hospitals: the Application of Security Safeguards. Acta Informatica Medica. 2016; 24(1):47-50.

[13] Mohammadzadeh N, Safdari R. Patient monitoring in mobile health: opportunities and challenges. Medical Archives. 2014; 68(1):57.

[14] Mohammadzadeh N, Safdari R, Rahimi A. Cancer care management through a mobile phone health approach: key considerations. Asian Pacific Journal of Cancer Prevention. 2013; 14(9):4961-4.

[15] Mehraeen E, Ghazisaeedi M, Farzi J, S M. Security Challenges in Healthcare Cloud Computing: A Systematic Review. Global Journal of Health Science. 2016;9(3):1-10.

[16] Schnall R, Mosley JP, Iribarren SJ, Bakken S, Carballo-Diéguez A, Brown W. Comparison of a User-Centered Design, Self-Management App to Existing mHealth Apps for Persons Living With HIV. JMIR mHealth and uHealth. 2015;3(3):e91.

[17] Aantjes CJ, Ramerman L, Bunders JFG. A systematic review of the literature on self-management interventions and discussion of their potential relevance for people living with HIV in sub-Saharan Africa. Patient Education and Counseling. 2014; 95:185–200.

[18] Tufts KA, Johnson KF, Shepherd JG, Lee JY, Bait Ajzoon MS, Mahan LB, et al. Novel interventions for HIV self-management in African American women: a systematic review of mHealth interventions. The Journal of the Association of Nurses in AIDS Care: JANAC. 2015;26(2):139-50.
[19] Nhavoto JA, Gronlund A, Chaquilla WP. SMSaude: Design, Development, and Implementation of a Remote/Mobile Patient Management System to Improve Retention in Care for HIV/AIDS and Tuberculosis Patients. JMIR Mhealth Uhealth. 2015; 3(1): e26.
[20] Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res. 2004: 10;6(4):e40.
[21] Nguyen HQ, Carrieri-Kohlman V, Rankin SH, Slaughter R, Stulbarg MS. Supporting cardiac recovery through eHealth technology. J Cardiovasc Nurs. 2004;19(3):200-208.
[22] Ramanathan N, Swendeman D, Comulada WS, Estrin D, Rotheram-Borus MJ. Identifying preferences for mobile health applications for self-monitoring and self-management: focus group findings from HIV-positive persons and young mothers. Int J Med Inform. 2013;82(4):e38-e46.
[23] Garofalo R, Kuhns LM, Hotton A, Johnson A, Muldoon A, Rice D. A randomized controlled trial of personalized text message reminders to promote medication adherence among HIV-positive adolescents and young adults. AIDS and Behavior. 2016; 20(5):1049-59.
[24] Swendeman D, Farmer S, Mindry D, Lee SJ, Medich M. HIV Care Providers' Attitudes regarding Mobile Phone Applications and Web-Based Dashboards to support Patient Self-Management and Care Coordination: Results from a Qualitative Feasibility Study. Journal of HIV and AIDS. 2016; 2(4): doi: http://dx.doi.org/10.16966/2380-5536.127
[25] Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Global health, science and practice. 2013;1(2):160-71.
[26] Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, et al. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. AIDS (London, England). 2011; 25(6):825-34.
[27] Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. Journal of medical Internet research. 2013;15(1):e1.
[28] Holstad M, Ofotokun I, Farber E. There's an app for that! Promoting antiretroviral medication adherence in rural Georgia: the music for health project. J Mob Technol Med. 2013;2(4S):18.
[29] Odeny TA, Bukusi EA, Cohen CR, et al. Texting improves testing: a randomized trial of two-way SMS to increase postpartum prevention of mother-to-child transmission retention and infant HIV testing. PloS one. 2014; 28:2307–2312.
[30] Himelhoch S, Mohr D, Maxfield J, Clayton S, Weber E, Medoff D, et al. Feasibility of telephone-based cognitive behavioral therapy targeting major depression among urban dwelling African-American people with co-occurring HIV. Psychology, health & medicine. 2011;16(2):156-65.
[31] Smillie K, Van Borek N, Abaki J, Pick N, Maan EJ, Friesen K, Graham R, Levine S, van der Kop ML, Lester RT, Murray M. A qualitative study investigating the use of a mobile phone short message service designed to improve HIV adherence and retention in care in Canada (WelTel BC1). Journal of the Association of Nurses in AIDS Care. 2014; 25(6):614-25.
[32] Aharonovich E, Greenstein E, O'Leary A, Johnston B, Seol SG, Hasin DS. HealthCall: technology-based extension of motivational interviewing to reduce non-injection drug use in HIV primary care patients–a pilot study. AIDS care. 2012;24(12):1461-9.
[33] Hasin DS, Aharonovich E, O'leary A, Greenstein E, Pavlicova M, Arunajadai S, Waxman R, Wainberg M, Helzer J, Johnston B. Reducing heavy drinking in HIV primary care: a randomized trial of brief intervention, with and without technological enhancement. Addiction. 2013;108(7):1230-40.
[34] Cornelius JB, Cato M, St Lawrence J, Boyer CB, Lightfoot M. Development and pretesting multimedia HIV-prevention text messages for mobile cell phone delivery. The Journal of the Association of Nurses in AIDS Care: JANAC. 2011; 22(5):407.
[35] Kirk GD, Himelhoch SS, Westergaard RP, Beckwith CG. Using Mobile Health Technology to Improve HIV Care for Persons Living with HIV and Substance Abuse. AIDS research and treatment. 2013;13(11):17-26.
[36] Forrest JI, Wiens M, Kanters S, Nsanzimana S, Lester RT, Mills EJ. Mobile health applications for HIV prevention and care in Africa. Current opinion in HIV and AIDS, 2015;10(6):464-71.
[37] Ingersoll K, Dillingham R, Reynolds G, Hettema J, Freeman J, Hosseinbor S, Winstead-Derlega C. Development of a personalized bidirectional text messaging tool for HIV adherence assessment and intervention among substance abusers. Journal of substance abuse treatment. 2014; 46(1):66-73.
[38] da Costa TM, Barbosa BJ, e Costa DA, Sigulem D, de Fátima Marin H, CasteloFilho A, Pisa IT. Results of a randomized controlled trial to assess the effects of a mobile SMS-based intervention on treatment

adherence in HIV/AIDS-infected Brazilian women and impressions and satisfaction with respect to incoming messages.International journal of medical informatics. 2012; 81(4):257-69.
[39] Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment & prevention: a systematic review of the literature. Open AIDS J. 2013; 7:17-41.
[40] Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, Jack W, Habyarimana J, Sadatsafavi M, Najafzadeh M, Marra CA. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. The Lancet. 2010; 376(9755):1838-45.
[41] Chang LW, Kagaayi J, Arem H, Nakigozi G, Ssempijja V, Serwadda D, Quinn TC, Gray RH, Bollinger RC, Reynolds SJ. Impact of a mHealth intervention for peer health workers on AIDS care in rural Uganda: a mixed methods evaluation of a cluster-randomized trial. AIDS and Behavior. 2011; 15(8):1776-84.
[42] Thomas B, Closson EF, Biello K, Menon S, Navakodi P, Dhanalakshmi A, Mayer KH, Safren SA, Mimiaga MJ. Development and open pilot trial of an HIV-prevention intervention integrating mobile-phone technology for male sex workers in Chennai, India. Archives of sexual behavior. 2015:1-2.
[43] Shet A, Arumugam K, Rodrigues R, Rajagopalan N, Shubha K, Raj T, D'souza G, De Costa A. Designing a mobile phone-based intervention to promote adherence to antiretroviral therapy in South India. AIDS and Behavior. 2010;14(3):716-20.
[44] Sharpe JD, et al. Interest in using mobile technology to help self-manage alcohol use among persons living with the human immunodeficiency virus: A Florida Cohort cross-sectional study. Substance Abuse. 2017; 19:1-6. doi: 10.1080/08897077.2017.1356793
[45] Shet A, De Costa A, Kumarasamy N, Rodrigues R, Rewari BB, Ashorn P, Eriksson B, Diwan V. Effect of mobile telephone reminders on treatment outcome in HIV: evidence from a randomised controlled trial in India. BMJ. 2014; 349:g5978 doi: 10.1136/bmj.g5978.
[46] L'Engle KL, Green K, Succop SM, Laar A, Wambugu S. Scaled-Up Mobile Phone Intervention for HIV Care and Treatment: Protocol for a Facility Randomized Controlled Trial. JMIR research protocols. 2015;4(1):e11.
[47] Kalichman SC, Kalichman MO, Cherry C, Swetzes C, Amaral CM, White D, Jones ML, Grebler T, Eaton L. Brief behavioral self-regulation counseling for HIV treatment adherence delivered by cell phone: an initial test of concept trial. AIDS patient care and STDs. 2011;25(5):303-10.
[48] Montoya JL, Wing D, Knight A, Moore DJ, Henry BL. Development of an mHealth Intervention (iSTEP) to Promote Physical Activity among People Living with HIV. Journal of the International Association of Providers of AIDS Care (JIAPAC). 2015: 2325957415601505.
[49] Miller CW, Himelhoch S. Acceptability of Mobile Phone Technology for Medication Adherence Interventions among HIV-Positive Patients at an Urban Clinic. AIDS research and treatment. 2013;2013:670525.
[50] Montoya JL, Georges S, Poquette A, Depp CA, Atkinson JH, Moore DJ. Refining a personalized mHealth intervention to promote medication adherence among HIV+ methamphetamine users. AIDS care. 2014; 26(12):1477-81.
[51] Niakan S, Mehraeen E, Noori T, Gozali E. Web and Mobile Based HIV Prevention and Intervention Programs Pros and Cons – A Review. Studies in Health Technology and Informmatics. 2017; 236:319-327.
[52] Mushamiri I, Luo C, Iiams-Hauser C, Ben Amor Y. Evaluation of the impact of a mobile health system on adherence to antenatal and postnatal care and prevention of mother-to-child transmission of HIV programs in Kenya. BMC Public Health. 2015;15(9):102-114.
[53] Mannell, J. Advancing gender equality to improve HIV prevention: A study of practice. African Journal of AIDS Research. 2016; 15(4), 315-323.
[54] Cooper V, Clatworthy J, Whetham J, EmERGE Consortium. mHealth Interventions To Support Self-Management In HIV: A Systematic Review. The open AIDS journal. 2017;11:119.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-180

Effects of Medical Device Regulations on the Development of Stand-Alone Medical Software: A Pilot Study

Kathrin BLAGEC[a,1], David JUNGWIRTH[b], Daniela HALUZA[b] and Matthias SAMWALD[a]
[a] *Center for Medical Statistics, Informatics, and Intelligent Systems. Medical University of Vienna*
[b] *Center for Public Health. Medical University of Vienna*

Abstract. *Background:* Medical device regulations which aim to ensure safety standards do not only apply to hardware devices but also to standalone medical software, e.g. mobile apps. *Objectives:* To explore the effects of these regulations on the development and distribution of medical standalone software. *Methods:* We invited a convenience sample of 130 domain experts to participate in an online survey about the impact of current regulations on the development and distribution of medical standalone software. *Results:* 21 respondents completed the questionnaire. Participants reported slight positive effects on usability, reliability, and data security of their products, whereas the ability to modify already deployed software and customization by end users were negatively impacted. The additional time and costs needed to go through the regulatory process were perceived as the greatest obstacles in developing and distributing medical software. *Conclusion:* Further research is needed to compare positive effects on software quality with negative impacts on market access and innovation. Strategies for avoiding over-regulation while still ensuring safety standards need to be devised.

Keywords. Medical Device Regulations, Clinical Decision Support Systems, Patient Safety

1. Introduction

Software that guides medical decision making has the potential to significantly transform and improve medical care. The increasing role of software in medical decision making also warrants caution about potential negative impacts experienced by end users. Such negative impacts can be caused by a wide variety of problems including errors or omissions in recommendations given to users, or by distracting and misleading users through usability issues. [1]

To address these issues, medical software products meeting certain criteria are covered by medical device regulations (MDR) under most jurisdictions. For example, in the USA, the distribution of medical devices is regulated by the U.S. Food & Drug Administration (FDA). [2] In the European Union, medical devices are regulated by Directive 93/42/EEC issued by the Council of the European Union. [3]

[1] Corresponding Author: Kathrin Blagec, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Währinger Straße 25a, 1090 Vienna, Austria. E-Mail: kathrin.blagec@meduniwien.ac.at

Importantly, these regulations also apply to 'standalone' medical software, i.e. software that is not connected to medical hardware, including mobile apps or web applications. [4] Thus, slight variations of the following heuristics are commonly used to determine whether such software is subject to MDR: (1) The software performs an action on data different from storage, archival, communication or simple search, (2) the performed action is for medical purposes, (3) the performed action is for the benefit of the individual patient. If all three points apply to the software in question it most likely qualifies as a medical device and is subject to MDR.

Previous research indicates that current regulations and standards might not be flexible enough to be applied to the growing sector of medical standalone software, such as mobile medical devices. [5] It is currently not clear how effective medical device regulations applied to software products are in improving software quality and safety. Furthermore, it is not known to what extent such regulations might also have negative effects, e.g., by obstructing the development of innovative medical software and decreasing its availability to medical professionals and patients.

In this pilot study, we explored software developers' perceptions of the ease of application of international medical device regulations and their views about the impact of these regulations on the development and distribution of standalone medical software.

2. Methods

We conducted an online survey among a convenience sample of domain experts with a past or ongoing involvement in a stand-alone software project where medical device regulations applied. Eligibility criteria required participants to have been involved in at least one software project where they have either (1) successfully achieved certification according to medical device regulations, (2) where they were actively working towards certification, (3) where MDR applied but where they had not yet actively worked towards certification, or (4) where they were planning to add functionality to their software so that MDR would apply.

Since medical software products are commonly distributed internationally, implicating that the respective products have to conform to multiple national and international regulations and legislations, we did not limit our target group to a specific country or region.

The survey consisted of three parts. In the first part, participants were asked about their roles in developing medical software. The second part encompassed 25 items using a 5-point Likert scale ranging from strongly agree to strongly disagree, asking respondents to indicate their level of agreement with several statements. These statements encompassed the impact of current medical device regulations on market access, software development processes, and the safety and quality of life of patients. The final part of the questionnaire consisted of four demographic questions.

Potential participants were identified through public registers of medical device associations, health start-up portals, app stores, references listed on websites of MDR consulting companies, academic research papers, and professional networks. In total, we identified 130 potential participants stemming from academia or industry, with the majority being located in European countries. Personal invitations were sent to all identified individuals in several waves between March 8 and April 5 2017.

3. Results

3.1. Characteristics of survey participants

In total, 23 participants completed the questionnaire, resulting in a response rate of 17.7%. Out of those, two participants had to be excluded because they did not meet inclusion criteria. Thus, the responses of 21 participants were considered for further analyses.

Demographic characteristics and experience with medical software development of survey participants are summarized in Tables 1 and 2, respectively. Ages ranged between

Table 1. Demographic characteristics of the survey participants.

	n	%
Gender		
Male	18	85.7%
Female	1	4.8%
Not stated	2	9.5%
Age		
20-39 years	9	42.9%
40-59 years	10	47.6%
60 years or older	1	4.8%
Not stated	1	4.8%
Residence		
Europe	19	90.5%
Asia	1	4.8%
Not stated	1	4.8%
Sector		
Academia	3	14.3%
Industry	18	85.7%
Main role in software project(s)		
Software engineering / software design / programming	4	19.0%
Middle or upper management	11	52.4%
Quality assurance	3	14.3%
Scientific research	1	4.8%
Regulatory / legal advice	2	9.5%
Years of experience working on software falling under MDR		
Less than a year	2	9.5%
At least one year but less than three years	6	28.6%
Three years or more	13	61.9%
Company / organization size		
Less than 10 employees	8	38.1%
10 - 49 employees	9	42.9%
50 - 249 employees	2	9.5%
Not applicable	2	9.5%
MDR relevant to participants' software projects		
Medical Device Regulations of the European Union or one of its members states	20	95.2%
Medical Device Regulations of the United States of America	4	19.0%
African Medical Device Regulations	1	4.8%
Asian Medical Device Regulations	1	4.8%
Types of software		
Mobile apps	10	47.6%
Desktop software	4	19.0%
Software embedded or linked with Electronic Health Record systems or Electronic Order Entry systems	8	38.1%
Web-based applications or services	12	57.1%
Other	4	19.0%
Which end-users did these software projects target?		
Medical professionals (e.g., medical doctors, nurses, pharmacists)	19	90.5%
Patients	7	33.3%

Table 2. Reported difficulties encountered in developing and distributing software falling under MDR, as well as various other questions. Underlined results indicate median responses.

	Strongly disagree	Disagree	Neutral / Don't know	Agree	Strongly agree
A) We found it difficult to develop and distribute software when medical device regulations applied because of...					
The uncertainty about whether medical device regulations applied (n=21)	3	6	2	8	2
The uncertainty about which medical device risk class applied to the software (n=21)	1	3	5	7	5
The additional time needed for development (n=21)	1	1	2	9	8
The additional costs (n=21)	0	1	3	8	9
The constraints imposed on the features of the software (n=21)	0	4	5	9	3
The constraints imposed on the software development process (n=21)	0	4	4	8	5
Our lack of expertise in regulatory matters (n=21)	1	2	7	7	4
Associated legal risks (n=21)	0	6	5	8	2
The heterogeneity of medical device regulations in different legislations (e.g., between the European Union and the United States) (n=21)	0	3	5	8	5
Frequent changes to medical device regulations (n=21)	1	9	5	5	1
B) Other questions					
At least initially, we found it difficult to judge if medical device regulations applied to our software (n=20)	0	8	1	6	5
When the customers are patients, they sufficiently recognize and reward certification so that it pays off to go through the regulatory process. (n=21)	2	7	10	1	1
When the customers are medical professionals (e.g., medical doctors, nurses, pharmacists), they sufficiently recognize and reward certification so that it pays off to go through the regulatory process. (n=21)	1	3	9	7	1
When the customers are institutional buyers (e.g., large health care organisations, hospitals), they sufficiently recognize and reward certification so that it pays off to go through the regulatory process. (n=21)	1	1	2	11	6

28 and 64 years with a median age of 40 years. Most respondents (n=19) were located in one of the European Union's member states. Consequently, for the majority of respondents, MDR of the European Union or one of its member states were relevant to their software projects.

Most respondents (n=14) had been involved in one or more software projects that successfully achieved certification according to medical device regulations. Five participants were involved in software projects where they were actively working towards certification. The remaining two respondents were either involved in software

Table 3. Reported effects of MDR on developed software products. Underlined results indicate median responses (where two responses are underlined, the median is between responses).

	Much worse	Worse	Slightly worse	No effect	Slightly better	Better	Much better	Don't know / Not applicable
Which effects did medical device regulations have on the following aspects of your software?								
Usability (n=21)	0	1	3	6	5	4	1	1
Reliability (n=21)	0	1	0	6	3	8	1	2
Protection of patients from harm (n=21)	0	1	0	6	6	4	3	1
Data security (n=21)	0	0	0	9	4	5	2	1
Performance (e.g., speed of execution, memory use) (n=21)	0	1	3	16	0	0	0	1
Feature richness (n=21)	0	2	5	11	1	1	0	1
Software maintainability (n=21)	2	1	4	6	5	1	0	2
Ability to change or add features when software has already been deployed (n=21)	4	5	3	6	0	0	1	2
Possibility of software customization / configuration by end users (n=21)	2	5	4	7	0	0	0	3

projects where MDR applied but where they had not yet actively worked towards certification, or they were planning to add functionality to their software so that MDR would apply. More than two thirds (n=16) of the participants had hired an external consulting organization to assist in the regulatory process in at least one project.

3.2. Difficulties with developing and distributing medical software

Reported difficulties with developing and distributing medical software caused by MDR are summarized in Table 2, section A. More than half of the respondents (n=11) agreed that they found it difficult to judge whether MDR applied to their software. The additional time and costs for regulatory processes were perceived as the greatest obstacles in developing medical software. The uncertainty about whether MDR apply, which risk class applies, lack of expertise in regulatory matters, heterogeneity of medical device regulations, constraints imposed on the software development process and on the features of the software were also perceived as barriers. However, associated legal risks and frequent changes to MDR played a lesser role.

We asked participants if they perceived that patients, medical professionals or institutional buyers (e.g. large health care organizations, hospitals) sufficiently recognize and reward certification so that it pays off for software developers to go through the regulatory process. On average, this was confirmed only for institutional buyers (Table 2, section B).

3.3. Effects of MDR on software products

Reported effects of MDR on developed software products are summarized in Table 3. On average, participants reported that going through the required certification process had slightly positive effects on the usability, reliability and data security of their software,

and only slightly improved protection of patients from harm. They reported no effect on software performance.

Slightly negative effects on feature richness and software maintainability were perceived by a small subset of participants. Most negative effects were reported for the ability to change or add features in already deployed software and the possibility of software customization and configuration by end users.

4. Discussion

4.1. Principal results

MDR aim to improve healthcare by ensuring that devices conform to safety standards and reflect the latest progress in science. Thus, they do not only apply to 'classical' medical devices such as pacemakers, insulin pumps, or in-vitro diagnostics, but also to medical standalone software. However, the requirements in the development and distribution of user-friendly and reliable medical standalone software may differ strongly from the development of more traditional medical devices, e.g. in terms of the need for customizability or ongoing updating and adding of new features. This raises the question whether current regulations and associated required certification processes are appropriate to duly serve their purpose in improving the quality of those products without unnecessarily impeding market access of potentially valuable software.

Our results indicate that the first challenge in developing and distributing medical standalone software lies in clarifying whether the software in question is at all subject to MDR. Within the past seven years, national and international regulatory agencies aimed to address this issue by publishing guidance documents and software classification schemes. These include the MEDDEV 2.1/6 published by the European Commission or the guidelines of the European Coordination Committee of the Radiological, Electromedical and Healthcare IT Industry (COCIR) and the International Medical Device Regulator Forums (IMDRF). [6,7] These documents can certainly help developers of medical standalone software by providing a rough guidance. Nevertheless, the various manifestations and wide range of applications of medical standalone software may not be fully covered by these documents. Thus, it is difficult to achieve legal certainty whether a specific software product needs certification, and to identify which risk class applies to the product in question.

Despite past and ongoing efforts of dedicated working groups such as the Global Harmonization Task Force (GHTF) [8] and its successor, the International Medical Device Regulators Forum (IMDRF) [8], to create international standards and thereby harmonize existing national and international MDR, the heterogeneity of regulations under different legislations still manifests as an obstacle in developing and bringing to market internationally competitive products.

Our results suggest that, overall, medical device regulations can help to slightly increase patient safety by improving certain aspects of software products. Further research is needed to compare these positive effects with the potentially negative impact current regulations and associated certification procedures may have on innovation in the health IT sector.

The process of developing and distributing medical software is very resource-intensive, both in terms of financial expenditure and time. It also imposes restrictions on the functionality of software products. These additional burdens may especially

discourage smaller-sized companies and startups from entering this market segment and consequently slow down innovation.

One solution to potential over-regulation proposed by Yang and Thompson is to implement a 'substantial dependence' standard to discern software that should be regulated from software that does not need regulation. [9] According to this standard, a software product that guides medical decision making would not need regulation if (1) the software is transparent in its data and recommendations, (2) the user is competent to interpret the recommendations, and (3) the user has adequate time to reflect on the recommendation.

4.2. Limitations

The convenience sampling approach in this small-scale pilot study somewhat limits the generalizability of our findings to the larger population of medical software developers. However, the results provided herein can serve as a foundation for follow-up qualitative or quantitative research exploring certain aspects, e.g., potential options to improve the current situation, in more detail.

The majority of respondents were employed at companies with less than 50 employees. Larger companies and organizations may have fewer difficulties complying with MDR, and are not well-reflected in the sample. Furthermore, the majority of our studies' respondents had already achieved certification in at least one project. Views and assessments of this sample may differ from views of those who abstained from entering this market segment because of the regulatory burden.

We did not seek information regarding the regulatory class of the software being developed, the main business of the organization responding, whether the organization developed within a quality management system (ISO 13485 or ISO 9001) or whether it worked to key medical software standards. As a result of a new risk classification system for medical software introduced by the new European MDR (2017/745) that came into force on 25 May 2017, more software will be assigned higher risk classes (II or III), thus being subject to even stricter requirements and scrutiny processes. Data collection of our study took place between March and April 2017, therefore the results presented herein only partially reflect the new regulatory situation. Future research should explore the implications of the new European MDR that will apply after a transitional period of three years (i.e. spring 2020) for the development and distribution of medical standalone software.

5. Acknowledgments

We would like to thank the survey participants who took the time from their busy schedules to participate in this study.

This project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement No 668353 (KB and MS).

6. References

[1] T. L. Lewis and J. C. Wyatt, "mHealth and Mobile Medical Apps: A Framework to Assess Risk and Promote Safer Use," Journal of Medical Internet Research, vol. 16, no. 9, p. e210, Sep. 2014.

[2] FDA, "Overview of Device Regulation." [Online]. Available: https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/Overview/ucm2005300.htm. [Accessed: 01-Jun-2017].
[3] Council of the European Union, Council Directive concerning medical devices 93/42/EEC. 1993.
[4] D. Luzi and F. Pecoraro, "Medical device software: a new challenge.," Stud Health Technol Inform, vol. 180, pp. 885–890, 2012.
[5] C. J. Vincent, G. Niezen, A. A. O'Kane, and K. Stawarz, "Can standards and regulations keep up with health technology?," JMIR Mhealth Uhealth, vol. 3, no. 2, p. e64, Jun. 2015.
[6] COCIR (European Coordination Committee of the Radiological, Electromedical and Healthcare It Industry), "COCIR Contribution: Decision diagram for qualification of software as medical device." 22-Nov-2010.
[7] European Commission, "MEDDEV 2.1/6: Guidelines on the qualification and classification of stand alone software used in healthcare within the regulatory framework of medical devices." 01-Jan-2012.
[8] "International Medical Device Regulators Forum." [Online]. Available: http://www.imdrf.org/. [Accessed: 01-Jun-2017].
[9] Y. T. Yang and B. M. Thompson, "Regulatory Framework for Clinical Decision Support Software: Present Uncertainty and Prospective Proposition.," J Am Coll Radiol, vol. 12, no. 7, pp. 672–675, Jul. 2015.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-188

Development of a Computer-Aided Dosage and Telemonitoring System for Patients Under Oral Anticoagulation Therapy

Heiko KRUMM[a,1], Nils REISS[b], Malte BURKERT[a], Thomas SCHMIDT[b], Steffen BIEHS[a], Christian BOHR[a], Fabian GÜRTLER[a], Hendrik HORN[a], Philipp KREUTZER[a], Philipp MEWES[a], Hendrik MILLER[a], Christian RIEST[a], Christian RÖMER[a], Albert SEEBOLD[a], Gertrude SPRUNG[a] and Oliver ZIEGLER[a]
[a]*Department for Computer Science, Technische Universität Dortmund, Germany*
[b]*Department for Clinical Research, Schüchtermann-Klinik Bad Rothenfelde, Germany*

Abstract. In this paper, we present a system that allows patients who require anticoagulation medicine an opportunity to independently manage their dosage concentration with the help of two machine learning algorithms. The basic idea is to predict the next dosage by using a neuronal network and the model predictive control approach, both based on the history of data already available from patients. This machine learning system is expanded by an smartphone application for the patients, and a website for the doctors to support their patients.

Keywords. artificial neural network, model predictive control, computer aided dosage, INR self management, anticoagulation therapy, telemedicine

1. Introduction

Today, due to heart disease and interventions, there are many patients who require anticoagulation medications (such as patients with mechanical heart valves, atrial fibrillation, LVAD-patients, etc. [1]). The correct dosage of the medication is vital: If the coagulability is too high, there is an increased risk of thrombosis. If it is too low, the risk of bleeding increases. To avoid these complications, it is important that the International Normalized Ratio (INR) remains in the therapeutic range [1,2]. To achieve this the medication dosage is usually determined by a doctor based on an INR-diary. However, due to the long intervals between visits to the doctor, the results in the aftercare were often not optimal. The INR-values therefore often lie outside the therapeutic range. Nowadays INR self-management has become established. Here, the patient measures his INR-value and determines his own medication dose. Several trials have shown the self-management to be beneficial both to therapeutic success and patient satisfaction, although there is room for improvement as deviations still occur regularly [2,3,4]. Another problem is the loss of overview for patients in self-management and overall control of the physician. To solve these problems our study aimed to develop a system that 1. automatically communicates the INR-specific values to the physician for documentation and control, and 2. creates an individually calculated medication recommendation for the patient based on personal health parameters and therapy relevant data records with the aim of optimizing the INR setting and to avoid complications.

[1] Corresponding author: Heiko Krumm, aDepartment for Computer Science, Technische Universität Dortmund, 44221 Drotmund, Germany. Email: krumm@cs.uni-dortmund.de

To illustrate the architecture of the system and the algorithmic methods in particular this paper is organized as follows. Section 2 presents an overview of related work and other approaches with the same goal. Section 3 presents the architecture of the system with focus on the two suggestion methods neural network and model predictive control. Tests and evaluation of our recommendation models are presented in section 4. A roundup and an outlook is offered in section 5.

2. Related Work

Heneghan et al. conducted a meta-analysis of randomized trials on self-monitoring INR-values and partly self-adjusted therapy using anticoagulant drugs [5]. They found that in the evaluated trials the self-management to improve the quality of oral anticoagulation with pooled estimates showing significant reductions in thromboembolic events, all-cause mortality and major hemorrhage. Ferreira et al. reviewed a telemonitoring system for VKA-patients and proved it to be safe and effective [6]. Poller et al. conducted a randomized study of two commercial computer-assisted dosage programs (PARMA 5 and DAWN AC), demonstrating the safety and effectiveness of these programs in comparison with experienced medical staff [7]. Both programs feature a graphical interface for patient management and monitoring for the medical staff, however only the DAWN AC-program contains the possibility of partial patient interaction, with the patient using a web browser to report INR values and receiving the dosage via text message. The algorithms used by PARMA 5 have been demonstrated by Manotti et al. [8]. The algorithms used in DAWN AC remain unknown. Rasmussen, Corell, Madsen and Overgaards investigated the management system CoaguTel [9]. The Hillingdon-algorithm, which is the main algorithm used by CoaguTel, has been described by Wilson and James [10]. None of the established and reviewed systems use a mobile application for patient interaction (e.g. sending INR-documentation), use a machine learning algorithm apart from regression analysis or consider patient vitamin K intake. We presume that because of this the existing systems lack the ability to easily incorporate new kinds of data records, which might be relevant to the dosage recommendation, such as the patient's dietary information.

3. Method, Experiment

The architecture of our system can be split into the two components front end and back end. The front end can further be split into a mobile app for the patients and a platform-independent web interface for the physician in charge. The back end is modularly developed with a restful web server, a database that stores the data of the patients and machine learning-based drug dosage recommendation systems [11]. For the patient's needs, a mobile app was developed for Android smart phones. This app is capable of tracking the INR measurements and other relevant data records, like a digital diary.

Upon entering a new data entry, the web server sends a request to two different machine learning systems based on the latest measurements. These two are an artificial neural network and a model predictive control system. The recommended drug dosage is stored in the database.

To monitor the recorded INR measurements and drug dosages, the physicians have access to a platform-independent web interface. Additionally, this interface informs the treating physician automatically, if a patient's record shows any threatening conditions (e.g. measured INR value is out of range).

To validate the system, 11 patients, all in VAK-treatment for of having a LVAD, have been asked to try out the system. Age range was 49 - 68. A total of 277 pairs of INR-measurements and dosages have been collected over a span of 12 weeks. INR target ranges were individual, but lied all between 2 and 3.

3.1. Assisted Patient Monitoring

Until now the patients had to keep the INR diary manually. This led to some disadvantages: While bad readability is one of the minor problems, the major disadvantage is that the patient visits his doctor in relatively large time intervals of two or three weeks, sometimes longer. This results in the fact that adjustments of the dosage can be applied only with a delay according to the changes of the measurements. Our app does not only replace the diary, but also transmits measured values to the doctor immediately. This way it is not only possible for the doctor to conduct adjustments of the dosage promptly but also to react to serious discrepancies of the INR values from the therapeutic range as soon as possible, optionally by adapting the dosage by using the web interface.

Some food and drugs affect the INR value. Unfortunately, there are almost no studies about this research field. It is only known which foods and drugs cause changes of the INR value. Therefore, it is additionally possible for the patient to document his food and drug history during his INR medication. The data collected can potentially be used in further work.

3.2. Automated suggestion on basis of data-driven machine learning algorithms

To suggest a dosage to the patient, we investigated two different machine learning techniques. First, we took a look at neural networks to predict the next dosage directly by learning from a set of training data records. Second, we build a control system, based on model predictive control, to influence the INR value to be in the middle of the TR[2]. This section provides the reader with the setup for the machine learning processes as well as an evaluation.

3.2.1. Neural Network Approach

Neural networks are a great approach, based on human's cells and neurons, to investigate a non-trivial connection between some data and a so-called label [12,13]. The advantage of neural networks consists of the ability to detect non-linear and complex correlations in the data. The net should predict a label for a data record only based on previous observations. A data record is mostly a set of values, with fixed size, and a label is a value associated with that data record. To use a neural network, a set of training data records is needed. A training data record contains the values and the corresponding correct label. After the network has been trained, it can be used for predicting a dosage. The training process itself is typically a gradient descent approach, minimizing the mean squared error between the (correct) label of a training data record and its prediction from the network.

The approach divides into the setup for the neural network and the (non-trivial) computation of a set of training data records. We used MATLAB's neural network

[2] TR: therapeutic range

toolbox[3] as the implementation of the neural network. There we used a simple feed-forward network with multiple layers of neurons. We extract the training data records from several sources. A data record contains information about the INR value, the dosage that was taken by the patient as well as a time stamp.

To train the network we created a set of about 700,000 training data records containing a fixed row of data entries, which consist of the last N data records of the patients INR and drug history. As the label, it contains the INR and dosage values for the next day. Here we suppose the weak assumption, that the patient knew, what was the best dosage at that time. Most patients in our data-stock were on a good adjustment, so that we suppose learning from them would create a good neural network for predicting the dosage. In addition to the last N data records from a patient, we took the type of medication (MARCUMAR or COUMADIN) and the TR (containing the maximum and minimum INR values) as input for our network. As a result, we get $2 * N + 3$ input neurons for a fixed $N \in \mathbb{N}$. The network has two output neurons for the dosage and the INR prediction.

Because of the non-deterministic random like behavior of neural networks, a secure mechanism should be implemented, which checks the predicted dosage against some kind of rule. We suggest only using dosage predictions in the "typical" range of a patient or check for heavy deviations in the dosage prediction.

3.2.2. Model Predictive Control

Another approach that is used to calculate a dosage recommendation for the patient is based on MPC (model predictive control). This method originates from control system theory and extends regular control systems with a model to predict the behavior of the system and to choose the optimal inputs. MPC based controllers take the current state of the system and optimize the system input under a given cost function. This optimization is calculated for a horizon of T steps so that at time t the cost-function J is minimized in the timespan $[t, t + T]$. Only the first step of the optimized control outputs is used and the calculation is started again [14,15].

In our case of finding an optimal dosage we interpret the patient's body as the system to be controlled. The dosage is used as the control input, while the measured INR is taken as the controlled variable. We use the center of the therapeutic range as the reference point for the control output so the cost function is given as the difference between this value and the INR (Eq. 1).

$$J = (TR_m - INR)^2 \tag{1}$$

This context requires the underlying model to represent the correlation between a given medication dosage and the resulting output, the INR value. We used a simple model which assumes an output that decreases over time and increases with the input. This simulates higher INR values with higher dosages and lower INR values with lower medication dosages.

Similar to the neural network approach the MPC structure was implemented with MATLAB and its *Model-Predictive-Control*-Toolbox[4]. The described model was realized with the help of Simulink, a software bundled with MATLAB to build, simulate and control complex systems.

[3] https://www.mathworks.com/products/neural-network.html
[4] https://www.mathworks.com/products/mpc.html

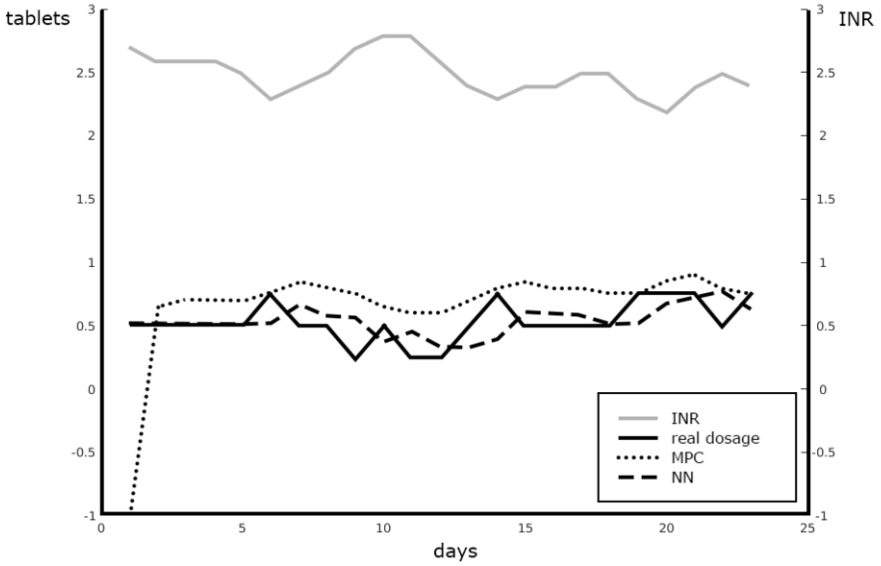


Figure 1. Comparison of different algorithms with real dosages and the corresponding INR values

Finally, both of these systems have to be triggered and the given dosages need to be transferred back to the server. This was realized with a message queue based infrastructure that took requests from the server and distributed them to the recommendation processes. Since both recommendation approaches were built with MATLAB a Java connector between the message queue and MATLAB was built. This connector reads the recommendation requests from the queue, starts up the recommender and reads their results. These results are then put into another queue from where the web server reads and persists them and finally displays them to the user in their app and the doctor in the web front end.

4. Results

Our app improves the process of INR self-management in several ways. The patient gets an easier way to keep track of his medication and INR history. The doctor is enabled to keep track of many patients at once and review their current state. Furthermore, they are notified in the event of INRs that are out of the TR. The modular infrastructure allows the evaluation of multiple models for dosage recommendation.

To evaluate the recommendation models their results were compared with actual dosages (Figure 1). Since these dosages – which were chosen by the patients themselves – are not necessarily optimal, a recommendation close to the actual dosage does not indicate a good recommendation. For a better assessment of the evaluation the patient histories we looked at 30 day blocks of measurements and classified them by variance of the INR in respect to the middle of the therapeutic range.

For the blocks with maximal, average and minimal variance the algorithms were compared to the actual dosages in the following way: Since the neural network needs at least a week of patient history prior to the evaluated day, a window of 7 days was moved along the 30-day block. This resulted in 23 different recommendations that could be

compared to the dosage value given in the dataset. The mean squared error E of these differences was then calculated (Eq. 2). These errors were then averaged over the different variance categories, as can be seen in Table 1.

$$E = \frac{1}{23} \sum_{n=8}^{30} \big(R(n-7,n) - D(n)\big)^2 \tag{2}$$

$R(m,n)$ = Recommendation for day n, with measurement input from day m to n
$D(n)$ = Actual dosage on day n

Table 1. Mean squared error between recommended and real dosage

Variance	MPC	NN
maximum	0.4515	0.4711
median	0.3954	0.0867
minimum	0.3835	0.0297

As already mentioned, the actual dosages can not be considered optimal. Therefore a heuristic was constructed to further evaluate the algorithms. Based on the time that a dosage of anticoagulant needs to take effect (given as about 2 days for MARCUMAR, cf. figure 2) the following assumptions were made:

When evaluating the dosage $D(n)$ at time n, we take the difference between this and the corresponding recommendation $R(n)$ as $\Delta_D = R(n) - D(n)$. We then look at the INR offset by the above mentioned span of 2 days $I(n+2)$ and compare this to the center of the therapeutic range TR, $\Delta_I = I(n+2) - TR$. If $\Delta_I > 0$, we can assume that the dosage has been too high and therefore we want Δ_D to be negative, meaning a lower recommendation than the actual dosage. Vice versa we want $\Delta_D > 0$ for $\Delta_I < 0$.

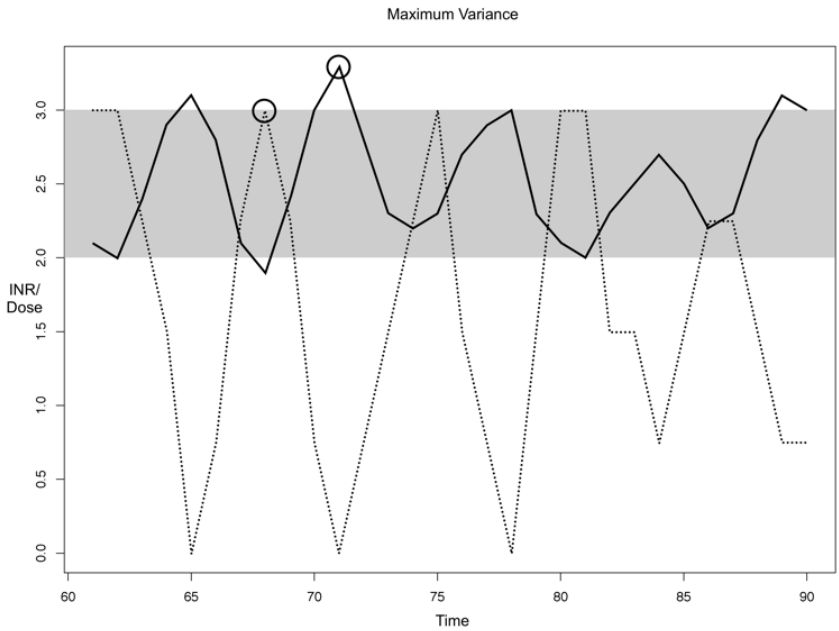
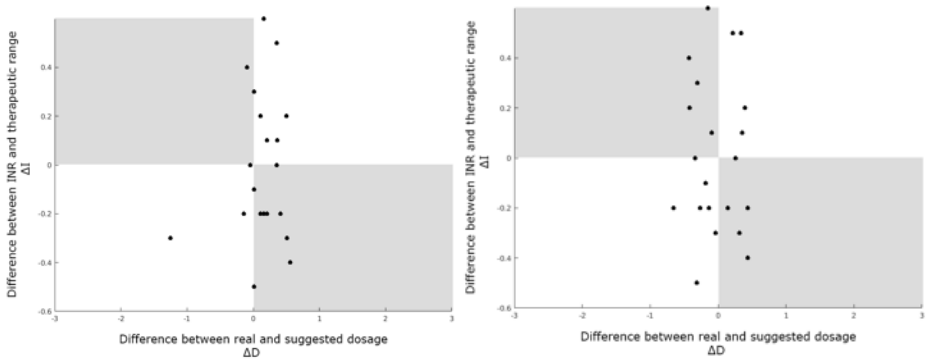


Figure 2. Patient history showing a high dosage (dashed) and the corresponding high INR-value 2 days later (solid)



(a) MPC (b) NN

Figure 3. Scatterplots of Δ_D and Δ_I, with the desired areas marked grey

An example plot of evaluating the two algorithms with this heuristic can be seen in figure 3 for one of our mini-trial patients with 29 data points. Other patients of our trial have similar distributions so that the algorithms have between 50% and 60% recommendations that meet the described criteria.

5. Discussion, Future Work

The mini-trial conducted at the *Schüchtermann* clinic showed the system to be viable in practical use. Our results support the idea that even relatively simple dosage recommendation models can be on a par with the patient's own dosage decisions and are in such consistent with previous assessments. The data records might not truly reflect the viability of the used dosage recommendation algorithms in practice as 9 out of 11 patients have been in close guidance by clinic caretakers during the trial period. A longer field trial with more completely self-monitoring participants is needed to fine-tune the algorithm parameters and determine a better quality estimation. Note that as of now while autonomous algorithmic decision-making for dosing is technically possible a human doctor must make the final decision by law in many countries (e.g. Germany).

The models have been trained on a very limited database. It can be assumed that they will perform better, when the training sets become bigger and of better quality. The integration of health tracking data, which in many cases is already being collected (like step count, other drugs, heart rate monitoring), is of particular interest. While the consideration of these data points by human caretakers is hardly viable, the machine learning algorithms can learn whether the integration is useful. The frameworks and standards already exist (e.g. *iOS Health app* and *ResearchKit* or *Google Fit*), a follow-up program would need to determine whether the data records can be used for dosage prediction. Besides automatically collected data records it also seems promising to include laboratory records, especially liver values and data records about the patient's diet, in the calculations. Some features, like detection of drug interactions, while theoretically useful, have been kept back as they were not necessary for the evaluation or were not feasible. The web application can be extended by an emergency access for fast insight in emergency situations. For practical usage, it should be possible to export the data points collected and calculated in the application to other programs

used in the clinic environment. In future, it would be desirable to develop an open sourced and publicly available system for automatic dosage recommendations to simplify comparisons among algorithms, as both researchers and patients would benefit.

Acknowledgements

We thank Prof. Dr. Sam Schulman of the McMaster University for his invaluable help in establishing an early dataset for training the models. Furthermore, we thank Prof. Dr. Nils Reiss, Petra Bartsch, Jan-Dirk Hoffmann and Thomas Schmidt of the Schüchtermann clinic in Bad Rothenfelde for their steady support and advice.

6. References

[1] E.P. Zeitler, Z.J. Eapen, Anticoagulation in Heart Failure: a Review, J Atr Fibrillation 8(1) (2015), 31-38.
[2] L. C Halder, L. B Richardson, R. Garberich, P. Zimbwa, M. K Bennett, Time in Therapeutic Range for Left Ventricular Assist Device Patients Anticoagulated With Warfarin: A Correlation to Clinical Outcomes, ASAIO Journal 63(1) (2016), 37–40.
[3] M. Pozzi, J. Mitchell, A.M. Henaine, N. Hanna, O. Safi, R. Henaine, International normalized ratio self-testing and self-management: improving patient outcomes, Vasc Health Risk Management 12 (2016), 387–392.
[4] L.J. Cromme, H. Voeller, F. Gaebler, A. Salzwedel, U. Taborski, Computer-aided dosage in oral anticoagulation therapy using phenprocoumon: Problems and approaches, Hämostaseologie 30(4) (2010), 183–189.
[5] C. Heneghan, P. Alonso-Coello, J. Garcia-Alamino, R. Perera, E. Meats, P. Glasziou, Self-monitoring of oral anticoagulation: a systematic review and meta-analysis, The Lancet 367 (9508) (2006), 404–411.
[6] F. Ferreira, E. Antunes, R.C. Neves, F. Farias, P. Malveiro, H. Choon, A. Galrinho, R. Cruz Ferreira, Telemonitorização de INR: Eficácia e Segurança de um Sistema de Avaliação em 453 Doentes, Acta Médica Portuguesa 25(5) (2012), 297-300.
[7] L. Poller, M. Keown, S. Ibrahim, G. Lowe, M. Moia, A. Turpie, C. Roberts, A. Van Den Besselaar, F. Van Der Meer, A. Tripodi et al., An international multicenter randomized study of computer-assisted oral anticoagulant dosage vs. medical staff dosage, Journal of Thrombosis and Haemostasis 6(6) (2008), 935–943.
[8] C. Manotti, M. Moia, G. Palareti, V. Pengo, L. Ria, A.G. Dettori, Effect of computer-aided management on the quality of treatment in anticoagulated patients: a prospective, randomized, multicenter trial of APROAT (Automated PRogram for Oral Anticoagulant Treatment), Haematologica 86(10) (2001),1060-1070.
[9] R.S. Rasmussen, P. Corell, P. Madsen, K. Overgaard, Effects of computer-assisted oral anticoagulant therapy, Thrombosis journal 10(1) (2012), 17.
[10] R. Wilson, A. James, Computer assisted management of warfarin treatment., Br Med J (Clin Res Ed) 289(6442) (1984), 422-424.
[11] A. Rodriguez, Restful web services: The basics, https://www.ibm.com/developerworks/library/ws-restful/index.html, last access: 7.3.2018.
[12] M.A. Nielsen, Neural Networks and Deep Learning, Determination Press, 2015.
[13] A. Abraham, Artificial Neural Networks, in: Handbook of Measuring System Design. John Wiley & Sons, 2005.
[14] J.B. Rawlings, Tutorial overview of model predictive control, IEEE Control Systems Magazine 20(3) (2000), 38-52.
[15] R. Dittmar, B.-M. Pfeiffer, Modellbasierte prädiktive Regelung - eine Einführung für Ingenieure, Oldenbourg Verlag, München, 2004.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-196

Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects

Antje WULFF[a,1], Birger HAARBRANDT[a] and Michael MARSCHOLLEK[a]
[a] *Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig and Hannover Medical School, Germany*

Abstract. *Background:* The availability of semantically-enriched and interoperable clinical information models is crucial for reusing once collected data across institutions like aspired in the German HiGHmed project. Funded by the Federal Ministry of Education and Research, this nationwide data infrastructure project adopts the openEHR approach for semantic modelling. Here, strong governance is required to define high-quality and reusable models. *Objectives:* Design of a clinical knowledge governance framework for openEHR modelling in cross-institutional settings like HiGHmed. *Methods:* Analysis of successful practices from international projects, published ideas on archetype governance and own modelling experiences as well as modelling of BPMN processes. *Results:* We designed a framework by presenting archetype variations, roles and responsibilities, IT support and modelling workflows. *Conclusion:* Our framework has great potential to make the openEHR modelling efforts manageable. Because practical experiences are rare, prospectively our work will be predestinated to evaluate the benefits of such structured governance approaches.

Keywords. Knowledge Management, Clinical Governance, openEHR, Health Information Interoperability

1. Introduction

The efficient reuse of once collected data can be described as one of the most pressing obstacles in the field of Medical Informatics. Current efforts not only focus on developing local solutions for data integration but address the challenges of cross-institutional data analytics and data sharing by proposing ideas for nationwide, interoperable data infrastructures. In Germany, a 120 million-worth funding initiative of the Federal Ministry of Education and Research (BMBF) was initiated to strengthen Medical Informatics [1]. As one of the funded projects, the HiGHmed consortia – consisting of three university medical centers in Heidelberg, Goettingen and Hannover – aspires at creating at least three Medical Data Integration Centers (MeDICs) based on a generic and scalable reference architecture for integrating data from care, research and external sources. HiGHmed perceives semantic interoperability as a prerequisite for enabling meaningful exchange of data in federated cross-institutional settings. Hence, much importance is attached to a semantically-enriched, interoperable and harmonized representation of data across institutions. To tackle this challenge, HiGHmed adopts the

[1] Corresponding Author: Antje Wulff, Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig and Hannover Medical School, Carl-Neuberg-Strasse 1, 30625 Hannover, Germany, E-Mail: antje.wulff@plri.de

openEHR approach for semantic modelling. Archetypes are used to define rich and computable metadata models of clinical information by applying constraints on a reference model [2].

As already stated by Garde et al. back in 2007 [3], the development of such clinical information models across institutional borders is challenging and requires close cooperation between stakeholders, data managers and leaders, as well as clearly defined processes and IT support. Strong governance, often referred to as 'Clinical Knowledge Governance' or 'Domain Knowledge Governance' [4], is needed to define high-quality, clinical relevant and reusable archetypes and templates on a cross-institutional level in a timely manner. Already back in 2006, the Australia's National E-Health Transition Authority (NEHTA) stated that 'undisciplined creation and application of archetypes threatens the goal of semantic interoperability' [5]. First practical experiences on implementing national governance schemes for archetypes in Norway support these theoretical considerations towards the urgent need of strict archetype governance [6]. Hence, for national infrastructure projects with the scope and extent of HiGHmed, a structured governance of archetype development and maintenance appears vital in order to "[...] achieve quality, reusability and interoperability in clinical models." [6]. With our work, we wanted to design a clinical knowledge governance framework for openEHR archetype modelling across institutions. This framework should help to optimize both cross-institutional and local modelling processes. Although we designed the framework specifically for HiGHmed, our work might serve interesting ideas for similar local, national or even international data infrastructure projects fostering secondary use of data.

2. Methods

Domain knowledge governance comprises "[...] all tasks related to establishing [...] formal and informal organizational mechanisms and structures in order to systematically influence the building, dissemination and maintaining of knowledge within and between domains" [7]. For learning on governance, we decided to visit European medical institutions active in developing health information exchange networks or national standardization programs. By conducting interviews with experts in Luxembourg (Luxembourg Centre for Systems Biomedicine), the Netherlands (Radboud University Medical Center), Norway (Bergen Nasjonal IKT), Denmark (Center for Innovative Medical Technology for the Odense University Hospital and the University of Southern Denmark), Slovenia (Ministry of Health and University of Ljubljana) and Austria (Medical University of Graz) we successfully identified practices as well as hurdles and pitfalls. Amongst others, we reviewed the results in terms of approaches for system sustainability, organizational structures and change management, use of state-of-the-art technologies, participant and stakeholder involvement and clinical modelling. We complemented the results with our own experiences on openEHR modelling and published work on archetype governance.

Much importance has been attached to the early involvement and recruitment of domain experts like clinicians [6]. In both the site visits and the published work, it became clear that *roles and responsibilities* – especially for model ownership and review activities – (e.g. a national editorial board) should be defined from the very beginning [3,6]. Furthermore, an efficient *IT support* enabling cooperation and collaboration in archetype designing and publication seems to be a well-known key success factor [6]. The definition of clinical information models on a face-to-face basis is cumbersome and

inefficient in nationwide and nearly impossible in international projects [4]. Hence, many of the current openEHR modelling activities over the world (including United Kingdom, New Zealand, Australia, Slovenia and Brazil) rely on efficient governance models backed by dedicated software tools supporting the governance, maintenance and publication of archetypes and related artefacts. A sufficient IT support that allows the collaborative authoring, commenting and reviewing of information models on a cross-institutional level is required. Well-known examples are tools like the web-based Clinical Knowledge Manager (CKM)[1], the LinkEHR Model Manager[2], ART-DECOR®[3] or Simplifier[4] .

A governance framework also should include fundamentals about *types and variations of archetypes* that might occur. As stated by Garde et al. (2007), there will be "[...] significant concept overlaps between the various health care domains [...]" [3] so that archetypes need to be standardized across all institutions to make them compatible. However, not all archetypes need to be standardized across fields, institutions or even nations [3]. Our own experiences on using the openEHR approach for integrating data from clinical application systems [8] also makes us aware of this problem: sometimes, some very specific 'support'-archetypes which are not worth standardizing across the community are needed to integrate data properly. Due to the considerations of Garde et al. [3,4] and our experiences, we decided to include our ideas on possible variations and their treatment in our governance framework.

We consider the design of workflows and processes for archetype modelling in HiGHmed as next step in designing a governance framework. Hereby, all responsibilities, relations and dependencies of tasks and communication needs between roles can be reconstructed. This includes the definition of workflows for specific events like first draft modelling, archetype specializations and reviewing or updating when knowledge changes [3]. For process modelling, we used the Business Process Modelling Notation (BPMN). Based on this approach and considerations, we were able to define an outline for our clinical knowledge governance framework that comprises content on IT support, types and variations of archetypes, roles and responsibilities, and archetype modelling processes.

3. Results

3.1. IT support

A sufficient IT support is needed to support the adoption of governance guidelines for archetype modelling. Because of our first satisfactory experience with the Clinical Knowledge Manager (CKM), we consider it as a tool supporting the governance, maintenance and publication of models and related semantic artefacts. Typical activities supported are requirements gathering, authoring, commenting, reviewing and balloting of information models. The CKM encourage the creation of work groups for each subdomain and provide discussion spaces to ease clinical concept evaluation. By incorporating functions for a rigorous artefact lifecycle management, the status of clinical models is made visible. On the example of the CKM, for HiGHmed one instance

[1] http://www.openehr.org/ckm/
[2] http://www.linkehr.com/
[3] https://www.art-decor.org
[4] https://simplifier.net/

might be available. To foster an efficient collaboration, we recommended dividing this HiGHmed CKM by the help of CKM projects and incubators. All resources which are important across institutions and use cases are presented at the *All Resource* section. We propose to create one incubator for each use case (e.g. Infection Control, Cardiology and Oncology) capsuling domain-specific clinical models. Furthermore, each institution gets its own CKM project for creating and discussing institution-specific models. By this approach, each institution owns its repository of clinical models (e.g. also for storage of archetypes not related to HiGHmed). Of course, archetypes of the overall CKM project and/or the international CKM still can be used within sub projects or incubators by referencing them.

3.2. Types and variations of archetypes

3.2.1. Archetypes important for all institutions and for all use cases

Managed in HiGHmed context by: HiGHmed Modelling Group (see chapter 2.3)
The archetypes of this category play a role in nearly every use case and institution. Archetypes might be promoted to this category when it appears that many use cases aim at designing the same clinical concepts. These archetypes have the potential to be uploaded to the global openEHR CKM and marked as 'published' in the same manner as current global archetypes (like *ACTION.procedure*, *OBSERVATION.blood_pressure*, *EVALUATION.problem_diagnosis*, *CLUSTER.symptom_sign*). When creating a CKM instance in HiGHmed, it is recommended to use these archetypes as a starter kit.

3.2.2. Archetypes important for all institutions but only for specific use cases

Managed in HiGHmed context by: Use Case Modelling Leader
There might be archetypes which are relevant for all institutions in the context of one use case. At this stage, no other use cases are interested in these archetypes. For structuring purposes, these archetypes will be uploaded and managed separately within so-called CKM incubators. Thus, users like domain experts won't be confused by many archetypes they are not interested in.

3.2.3. Archetypes important for one institution

Managed in HiGHmed context by: Local Use case Chief Data Steward
When using the openEHR-based approach for data integration, some 'support'-archetypes are needed because of primary source systems. These archetypes won't be designed as a 'maximum set' of clinical concepts like encouraged by openEHR. However, although they won't bring an added value for the overall openEHR global community, they are crucial for integrating specific data sets in the HiGHmed context.

3.3. Modelling roles, responsibilities and tasks

Well-defined roles, responsibilities and tasks are vital for realizing structured modelling processes as well as for avoiding redundant modelling activities. Figure 1 shows the proposed modelling roles and groups for HiGHmed. Each role can be translated into one of the CKM member roles *Editor* and *Reviewer*.

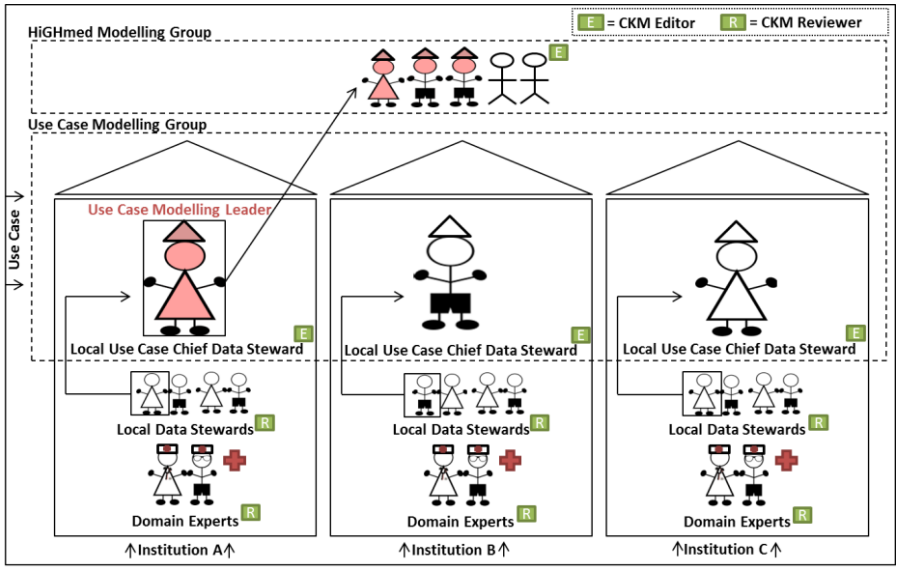


Figure 1. Recommended modelling roles and groups in HiGHmed on the example of three institutions

The superordinate **HiGHmed Modelling Group** (CKM Editor) comprises the leading modelers of the core use cases and modelling experts which are not related to specific use cases. The group should consist of a well-balanced mixture of members from different institutions and use cases. The members should have access to all incubators and use case activities. The core tasks of this group are management of the CKM project including the set-up of initial structures with an archetype starter kit. They lead the modelling and ongoing management of archetypes which are not exclusively related to a specific institution and/or use case (e.g. CKM check outs, change requests, uploads, revisions, invitations, review rounds). This group communicates with the global openEHR community and observes the global openEHR archetype repository.

Every institution should have **Local Data Stewards** (CKM Reviewer) who are responsible for the communication with domain experts to gather information about local requirements. With respects to the current HiGHmed personnel planning, these positions often will be taken over by clinical staff. Local Data Stewards with more technical backgrounds will be available directly out of the Medical Data Integration Centers (MeDICs). The Local Data Stewards don't need to be linked to a specific use case. However, every institution should declare one of the MeDIC-Data Stewards to a **Local Use Case Chief Data Steward** (CKM Editor). Local Data Stewards are responsible for analysis of requirements and local systems and for CKM side activities like administration of the resource center and translations. Moreover, they create first drafts for archetypes and participate in review rounds. They also should be capable of supporting the data integration specialists. The Local Use Case Chief Data Steward is responsible for managing the institution-specific CKM project and uploading archetypes which are exclusively needed locally. He/she leads and manages the Local Data Stewards and their participation in use case specific CKM check outs, change requests, uploads, revisions, invitations and review rounds. Also, he/she requests requirements analyses and forwards the results to the Use Case Modelling Leader. In close cooperation, they model first archetypes and design templates.

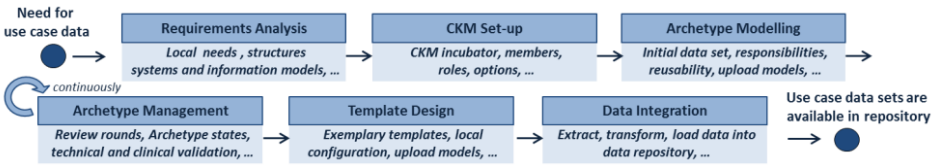


Figure 2. Executive overview of the governance process for modelling in HiGHmed

For each use case, a **Use Case Modelling Group** consisting of all Local Use Case Chief Data Stewards will be formed. This group defines one of its members as Use Case Modelling Leader who is responsible for the management of the cross-institutional but use case specific modelling activities. As **Use Case Modelling Leader** (CKM Editor), he/she will participate in the HiGHmed Modelling Group. He/she takes over the lead and management of the Local Use Case Chief Data Steward, the summarizing of local requirements, the definition of relevant clinical concepts and the realization or delegation of modelling tasks. In CKM, he/she creates the CKM use case incubator and its initial structure and coordinates CKM check outs, change requests, uploads, revisions, invitations and review rounds for use case specific archetypes.

The availability of **Domain Experts** (CKM Reviewer), e.g. clinicians or technicians, is crucial for an adequate requirements analysis, the modelling of first drafts and the review of archetypes by providing domain or technical knowledge. Some of the Local Data Stewards might be clinicians and, thus, domain experts. However, ideally, there are other experts which are neither part of the Use Case Modelling Group nor the HiGHmed Modelling Group.

3.4. Collaborative and coordinated modelling approach

The allocation, relations and dependencies of tasks and communication needs between roles are presented as archetype modelling workflow (uploaded on ResearchGate[1]).

A high-level view of the process is presented in Figure 2. The process starts by any need to standardize data items and reuse them. The needs will be determined by the selected use cases and by the input of stakeholders (like researchers, clinicians or industry partners). The need will be transferred to the Data Stewards so that new Use Case Chief Data Stewards and cross-institutional Use Case Modelling Groups can be set up. The HiGHmed Modelling Group will be informed about the new use case and its Use Case Modelling Leader. Simultaneously, a use case CKM incubator will be set up. Then, the leader will request for requirements of institution-specific needs, structures and systems by asking the Use Case Chief Data Stewards. Local Data Stewards will work closely with the Local Domain Experts to gather information. The analysis results will be documented by creating mind maps or first draft archetypes. After the requirements of all institutions will have been gathered, the actual modelling process will start. In close cooperation with the Use Case Modelling Group, the Use Case Modelling Leader will combine the local analysis results and will work out relevant clinical concepts. Then, it will be checked whether these concepts are exclusively needed at one institution only. In that case, the corresponding Local Use Case Chief Data Steward will take over the ongoing design and management of the archetype. The continuous monitoring of such cases as well as the proposal of specifications for institution-specific archetypes instead

[1] https://www.researchgate.net/publication/322519762_Archetype_Modelling_Workflow

of the development of new archetypes are core tasks of the Use Case Modelling Leader. The modelling process will be continued by searching in the HiGHmed CKM or the global CKM for suitable archetypes that already exist. If available, the archetype will be uploaded to the incubator as 'referenced archetype'. If not, a first draft archetype is created de novo. This means, that an archetype class as well as appropriate elements, entry features and terminology bindings will be created. At this step, it has to be checked whether other standard definitions (e.g. by HL7) can be used for archetype design. The archetype will be uploaded to the incubator or the institution-specific CKM project. When a new first draft will have been uploaded, the HiGHmed Modelling Group will be automatically informed by the CKM in order to check for redundant concepts and/or archetypes across institutions and use cases. If a redundant archetype occurs, the HiGHmed Modelling Group will take over the continuous management of this archetype from the Use Case Modelling Leader. The archetype will be included into the HiGHmed CKM project space and will change to a 'reference archetype' within the incubators.

The sub process *Manage Archetype* summarizes the most important activities throughout an archetype lifecycle. Here, the initiation of review rounds are of particular importance because through this the Local Data Stewards and Domain Experts will be empowered "[...] to create and change the knowledge inherent in archetypes, thus controlling the way EHRs [Electronic Health Records] are built up using designed structures to express the required clinical data [...]" [3]. It is recommended to reduce the number of face-to-face meetings between domain experts of different institutions as far as possible, e.g. only for setting up general goals of their use case. Any other discussions should be held within the CKM. After the archetype will have been approved, the Local Use Case Chief Data Stewards will start to design und upload templates by using the formerly approved archetypes. Exemplary templates can be discussed within the Use Case Modelling Group and reused across institutions but, in general, templates should be designed and uploaded within an institution-specific space.

4. Discussion

Our work contributes a novel framework for clinical knowledge governance in nationwide data infrastructure projects like HiGHmed. By analyzing related work on clinical knowledge governance as well as by visiting different sites and using their experiences on standardization projects (openEHR, FHIR, CDA, EDIFACT and others) –, we were able to carefully work out the key aspects that need to be covered in such a framework. We successfully gathered potential archetype variations, proposed roles and responsibilities and IT support, and outlined a workflow for archetype creation and approval.

For HiGHmed, the availability of high-quality and reusable archetypes is crucial for the achievement of the superordinate project objectives. By reusing once collected data, new and high-performing solutions for medical data analytics can be developed in order to accentuate the benefits of increased digitalization in medicine for patients, clinicians and researchers. As a prerequisite, data has to be integrated safely, accurate and semantically enriched to ensure the correct interpretation by humans and machines across institutions. Due to the inherent complexity of the clinical domain, modelling processes can be time-consuming. In our opinion, these efforts are manageable when following a structured modelling approach as proposed. By defining clear responsibilities and workflows not only the quality of final archetypes can be increased

but the development time of archetypes can possibly be decreased. Additionally, the initial modelling efforts might decrease in future as soon as a high-quality stack of published and reusable archetypes is available.

We agree with Garde et al. [3,4] and Bakke [6] that IT support plays a key role in governance. We appreciate that a major part of archetype governance can be covered by sufficient tools like the CKM. It has been reported that the use of the CKM facilitates and encourages the involvement of clinicians and domain experts in clinical information modelling [6]. For other standardization efforts similar tools for supporting collaborative work on models are available (e.g. ART-DECOR® provides template editors for HL7 V3/CDA templates, a value set editor and a terminology browser, lifecycle management features and others). These tools are also helpful for our approach because existing definitions can be searched and used as information input for archetype design. However, because HiGHmed adopts the openEHR approach for semantic modelling, we are in need of an openEHR-based governance tool. Furthermore, we think that such tools only can be used efficiently when modelling rules, roles and responsibilities – regardless of the standard or tool decision – are clearly defined and communicated beforehand. Hence, we decided to put more effort into governance than just implementing IT support. Because everyone has to understand the centerpiece of our framework – the modelling approach – we set up an internal modelling and governance workshop in HiGHmed.

We are aware that our work lacks in practical evaluation of the clinical governance framework. Currently, the long-term establishment and practicability of our approach is not assessable yet but we strive at implementing and optimizing the presented framework within the HiGHmed project as well as publishing evaluation results. We will define criteria to evaluate the impact of the framework. Important measures might be the time to archetype publishing as well as the model quality. Furthermore, the amount of redundant models, the transparency and sharing of models and the rate of reused archetypes with and without such a framework could be meaningful measures. Overall, we see a great opportunity to evaluate the use of clinical governance frameworks in nationwide projects of such extent for the first time.

5. References

[1] Bundesministerium für Bildung und Forschung, Digitalisierung in der Medizin: Medizininformatik, https://www.bmbf.de/de/medizininformatik-3342.html, last access: 10.1.2018.
[2] The HiGHmed Consortium, HiGHmed, http://highmed.org/, last access: 10.1.2018.
[3] S. Garde, P. Knaup, E. Hovenga, S. Heard, Towards semantic interoperability for electronic health records, Methods Inf Med. 46(3) (2007), 332–343.
[4] S. Garde, Clinical knowledge governance: The international perspective. Stud Health Technol Inform. 193 (2013), 269–281.
[5] National E-Health Transition Authority, Review of Shared Electronic Health Record Standards - Version 1.0 (2016).
[6] S. Ljosland Bakke, National governance of archetypes in Norway. Stud Health Technol Inform. 216 (2015), 1091.
[7] S. Garde, S. Heard, E. Hovenga, Archetypes in Electronic Health Records: Making the case and showing the path for domain knowledge governance, in: HIC 2005: 13th Australian Health Informatics Conference. H. Grain and M. Wise M, editors. Health Informatics Society of Australia, Melbourne Brunswick East, 2005.
[8] B. Haarbrandt, E. Tute, M. Marschollek, Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository, Journal of Biomedical Informatics, 63 (2016), 277-294.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-204

Modeling of ETL-Processes and Processed Information in Clinical Data Warehousing

Erik TUTE[a,1], Jochen STEINER[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of the University of Braunschweig - Institute of Technology and Hannover Medical School, Germany*

Abstract. Background: Literature describes a big potential for reuse of clinical patient data. A clinical data warehouse (CDWH) is a means for that. Objectives: To support management and maintenance of processes *extracting*, *transforming* and *loading* (ETL) data into CDWHs as well as to ease reuse of metadata between regular IT-management, CDWH and secondary data users by providing a modeling approach. Methods: Expert survey and literature review to find requirements and existing modeling techniques. An ETL-modeling-technique was developed extending existing modeling techniques. Evaluation by exemplarily modeling existing ETL-process and a second expert survey. Results: Nine experts participated in the first survey. Literature review yielded 15 included publications. Six existing modeling techniques were identified. A modeling technique extending 3LGM² and combining it with openEHR information models was developed and evaluated. Seven experts participated in the evaluation. Conclusion: The developed approach can help in management and maintenance of ETL-processes and could serve as interface between regular IT-management, CDWH and secondary data users.

Keywords. Data Warehousing, Organizational Models, Health Information Interoperability

1. Introduction

Resulting from the broad adaption of information technology (IT) in hospitals, there is much clinical patient data digitally available. The use of such data for other purposes than the originally indented is commonly referred to as its secondary use or reuse [1]. The Medical Informatics literature describes a big potential for reuse of this data, e.g. [1]. There are still technical and organizational challenges related to secondary use of clinical data [1]. A clinical data warehouse (CDWH) is a facility addressing some of these challenges by integrating data from heterogeneous sources and making it available for reuse, e.g. for analysis. The process of *Extracting* data from a source system (e.g. a patient data management system), *Transforming* it into a target schema (for example a schema facilitating later analyses) and finally *Loading* it into the persistence layer of the CDWH (layer in which data is stored permanently for later reuse) is called ETL-process. These ETL-processes are typically complex and custom-made which makes their management and maintenance challenging. Established tools supporting the graphical user interface (GUI) based development and operation of ETL-processes exist (e.g. Microsoft SQL Server Integration Services, Talend Open Studio). However, to the best of the author's knowledge there is a lack of established standards for the modeling of

[1] Corresponding Author: Erik Tute, Peter L. Reichertz Institute for Medical Informatics of the University of Braunschweig - Institute of Technology and Hannover Medical School, Mühlenpfordtstr. 23, 38106 Braunschweig, E-Mail: erik.tute@plri.de

ETL-processes and the thereby processed information, which are suitable for supporting ETL-process documentation and IT-management.

Modeling techniques and tools aiming to support IT-management in hospitals exist for more than 20 years. Their aim is to provide information or to give the possibility for analysis of business processes, information objects, supporting application systems and underlying hardware [2]. An approach to interoperability between different healthcare information systems through shared implementable definitions of the clinical concepts represented by clinical data are clinical information models (CIM), also referred to as detailed clinical models [3]. CIM-based CDWHs have potential merits ([4], [5]). This work is based on the assumption that there is an overlap between the information needs of regular hospital IT-management, ETL-development/maintenance and users in data reuse scenarios justifying the effort for standardized and machine processable documentation of ETL-related metadata. Figure 1 shows this information overlap without trying to be exhaustive. Metadata needs are derived from Kahn's proposed recommendations on metadata reporting for distributed data networks [6].

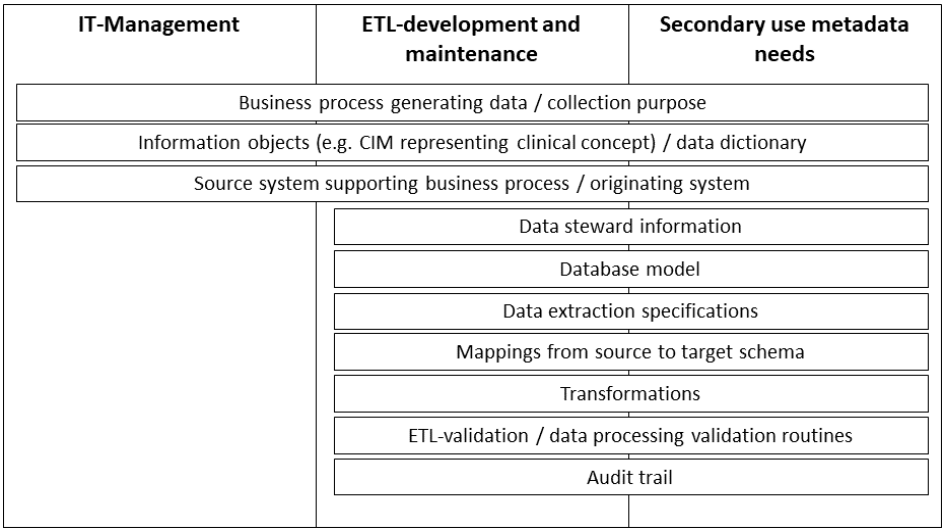


Figure 1. Information overlap between IT-management, ETL-development and secondary use.

Objective of this work is to find an approach for modeling of ETL-processes and thereby processed information, which can be integrated with tools supporting regular IT-management, can help in management and maintenance of ETL-processes and eases the machine processable provision of IT-management and ETL-related metadata for secondary use.

2. Methods

First, an expert survey and a literature review were conducted to find requirements for the modeling approach and to assess existing modeling techniques. Subsequently, an approach for modeling of ETL-processes for clinical data warehousing was developed as extension of an existing modeling technique. Finally, this approach was evaluated through a proof of concept (POC) modeling example and another expert survey.

2.1. Literature review on existing modeling techniques

In a literature review Pubmed, IEEE Explore and Scopus where searched for publications on modeling of ETL-processes. The search terms where developed iteratively for optimizing precision and recall. One reviewer included or excluded results based on previously defined criteria assessing first title, in doubt abstract and if necessary full text. References of included papers were also assessed. The final search terms as well as inclusion and exclusion criteria can be found in [7].

2.2. Expert survey on existing modeling techniques and their shortcomings

A survey employing face-to-face interviews or an online questionnaire depending on the experts' availability was conducted to complement existing ETL-process modeling techniques as well as to find shortcomings of these existing methods and currently practiced ETL-process documentation (for interview structure and online questionnaire see [7]). Experts were selected from the clinical data warehousing staff at Hannover Medical School (MHH) and from members of the HiGHmed consortium. HiGHmed is funded by the Federal Ministry of Education and Research (BMBF) under the Medical Informatics funding scheme, pursuing among others the objective to develop information infrastructures to increase efficiency of clinical research. These infrastructures are facing the problem of ETL-process documentation and metadata provision in a multisite environment [8]. Experts were recruited from all positions (e.g. ETL-programmers, data analysts, management) to gain different views on ETL-modeling and related information needs. Part of the interviews was a self-assessment of the presumed experts on their knowledge about ETL-processes.

2.3. Determination of ETL-modeling technique

Based on literature review and interview results, requirements for modeling of ETL-processes in clinical data warehousing were derived and prioritized as must-, should- or can-criteria. In a further step, evaluation criteria for the found existing modeling techniques where derived under consideration of their relevance for the objectives of the underlying master's thesis [7]. For each found modeling technique and each evaluation criterion an assessment regarding satisfaction of the criterion including documentation of the corresponding reasoning was done.

A definition of the modeling process based on the most promising modeling technique was developed. The intention in this step was to keep the changes or extensions in the modeling technique as little as possible in order to stay compatible with existing tools and not to distort core concepts of the modeling technique.

2.4. Evaluation of modeling technique

As a POC an ETL-process from the MHH clinical data warehouse was modeled using the modeling approach determined in the previous steps. Criteria for the selection of the ETL-process for POC where pragmatic: access to the actual ETL-process and respective documentation, good reachability of the ETL-developer and already available CIMs.

Another expert survey on the resulting ETL-model from the POC, the determined ETL-modeling approach in general and potentials for improvements and further development of the modeling approach complemented the evaluation. The survey

consisted of a short presentation of the modeling approach and the POC results for the experts who participated in the first survey in face-to-face interviews followed immediately by a paper-based questionnaire (see [7]). The presentation was given in groups, the questionnaire was completed by each expert on their own. Criteria for evaluation were shortcomings identified in the first expert survey. Improvements on these shortcomings were assessed based on experts' responses in the second survey.

3. Results

In the following, we briefly cover the results of the literature review and expert survey. We present the developed modeling technique extending 3LGM² in combination with openEHR information models and conclude with a description of the evaluation results. More details can be found in [7].

3.1. Literature Review and expert survey

The search in Scopus listed 299 publications, IEEE listed 420 and PubMed 13. From these 433 items, 76 were removed as duplications. After application of exclusion criteria on title and abstract, 55 publications remained for screening of the full text. Based on the references of not excluded full texts another 25 full texts were screened. Finally, 15 publications were included and found ETL-modeling techniques were documented.

Seven experts participated in face-to-face interviews. One interview was excluded afterwards, because the interviewed expert's self-assessment on ETL-knowledge indicated a lack of knowledge on ETL-processes. Two experts participated via the online questionnaire. One of those questionnaires was not completed and thus excluded.

3.2. Existing ETL-modeling techniques

Literature review and expert survey identified six existing ETL-modeling techniques (origin from review or survey specified in brackets):

- UML-based modeling of ETL-processes described in [9] and [10] (review)
- UML-based modeling of ETL-processes employing activity diagrams described in [11]. This method builds on reasoning from [9] (review)
- Arktos – Graph-based modeling of ETL-processes described in [12] (review)
- KoMo – conceptual modeling of ETL-processes described in [13] (review)
- Business Process Model and Notation (BPMN 2.0) based modeling of ETL-processes described in [14] (review)
- A combination of an extension of the 3LGM² technique for modeling of hospital information systems described in [2] with standardized CIMs (survey)

The only modeling technique resulting from the expert survey employs 3LGM² because it was already in use in medical informatics research and teaching contexts at MHH. The idea to combine 3LGM² with a standard able to express implementable definitions of clinical concepts resulted from the information overlap depicted in Figure 1 and aims at increasing reuse of content definitions in different contexts.

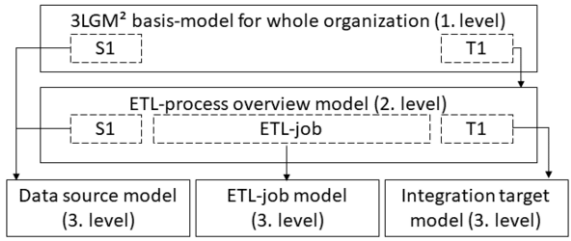


Figure 2. General structure of ETL-process models. S = Source, T = Target, Dashed-boxes: elements linking to sub-models, solid boxes: sub-models, arrows: links to sub-models. This is a modified version of a figure from [7].

3.3. ETL-modeling technique

The assessment regarding satisfaction of the criteria derived from the expert survey lead to the selection of the approach employing an extension of 3LGM². 3LGM² was combined with CIMs complying with openEHR [15], because the Peter L. Reichertz Institute for Medical Informatics at the MHH already has research activities regarding openEHR in clinical data warehousing (for example CIMs see [16]). However, other standards for CIMs could also be employed.

First development result was a mapping of required ETL-process components (e.g. data source, data input description, transformations, error-output description, data output description, integration target) into the 3LGM²-modeling technique. Second result was a reference model for ETL-process modeling with 3LGM² and openEHR defining required components and information to be stored in an ETL-process model.

Figure 2 depicts which kind of information is located in which levels of the model. The first level is a regular 3LGM² model for an organization supporting IT-management in general (for an example view on 3LGM² see [17]).

The second level sub-model gives an overview of an ETL-job by providing information on an ETL-processes' data sources and integration targets. An example view is shown in Figure 3.

On the third level, sub-models describing the data sources (including a description of the data input from the source like tables, fields, datatypes, constraints or a CIM), ETL-job (transformations, error-output) and integration target (including a description of the data output by linking a CIM) are located. Figure 4 shows an example view for an ETL-job. Sub-models can be linked in models of higher level, supporting reuse of information on the different levels and thus, improving maintainability.

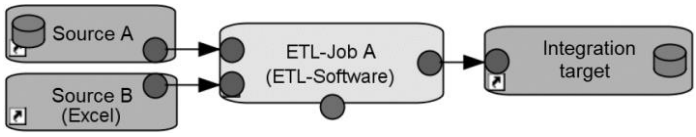


Figure 3. Example view for the ETL-process overview sub-model (level 2). Round edged boxes: application systems, circles: interfaces, arrows: indicate directed data flow, small boxes with arrows: indicate links, e.g. to sub-model or CIM, cylinder: database.

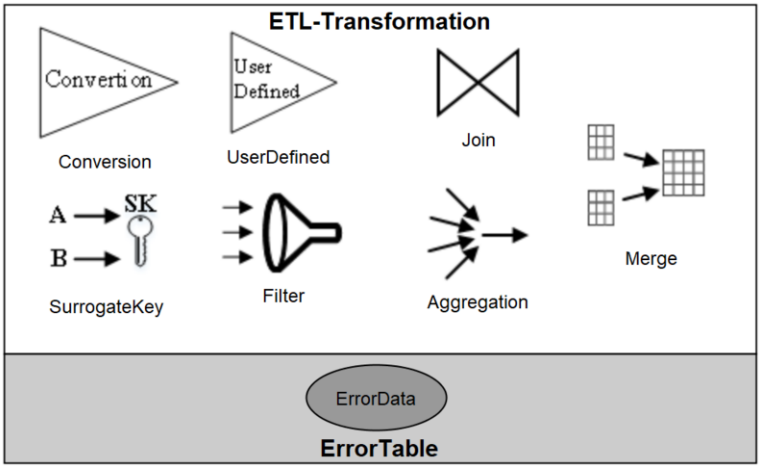


Figure 4. Example view for the ETL-job sub-model (level 3).

3.4. Evaluation results

A POC using the example of an ETL-job integrating assisted ventilation data from two different patient data management systems (PDMS) of a pediatric intensive care unit at the MHH into a common data model of the local clinical data warehouse revealed no issues.

All seven experts who participated in the face-to-face interviews participated also in the evaluation. Experts' answers showed improvements regarding the shortcomings identified in the first expert survey. These shortcomings were (a) missing or incomplete documentation of ETL-processes, (b) too much time necessary to understand ETL-processes implemented by others and (c) data quality issues. On (a): Three of seven experts answered "yes" to the question if they thought that the modeling technique could resolve the problem of no or incomplete ETL-process-documentation. Three experts answered "no" on that question and one expert did answer neither "yes" nor "no". Two of the three experts answering "no" motivated their answer stating that the approach first has to prove its applicability in practice. On (b): Four of seven experts answered "yes" to the question if the proposed modeling technique could resolve the problem of too much time necessary to understand unknown ETL-processes. Two of the three answering "no" to that question stated that the solution could not resolve the problem, but would reduce the burden. On (c): Five of seven experts answered "yes" to the question if they thought that this modeling technique could contribute to an increased data quality.

The experts' answers also indicated potential for further improvements of the modeling technique. The experts mentioned:

- More elements for describing transformations (e.g. union, pivot, sort)
- More details on transformations and better integration with actual ETL-job, e.g. by implementing automatic information extraction from ETL-job files from common ETL-tools
- Functional improvements of supporting tools, e.g. regarding model analysis or merges of models

4. Discussion

Main result of this work is an approach for modeling of ETL-processes and thereby processed information based on 3LGM² and openEHR. The modeling technique can be applied utilizing existing tools. Evaluation with experts from the intended user groups indicate positive expectations regarding its application.

The approach utilizes existing standards originally intended to support regular IT-management and interoperability of routine IT-systems. Thus, there is potential for reuse of common information items from routine IT-management and clinical data warehousing. Such reuse could reduce the efforts necessary to curate models on both sides regular IT-management and clinical data warehousing. Furthermore, there is a potential to reduce a bottleneck experienced in clinical data warehousing which results from the dependency on knowledge and support from experts not explicitly participating in data warehousing, e.g. an operator of a source system [18]. Another potential advantage of the resulting models is that these could be a suitable basis for the automated extraction of metadata in secondary use scenarios as described in the introduction.

As some experts stated in the evaluation an obvious limitation of this work is that we cannot provide sound results on real world application of the presented approach yet. As a result, the potential advantages mentioned above are uncertain and practical issues hampering the adoption of long-known modeling techniques like 3LGM² [2] until today may remain predominant.

Despite of the good intention not to distort core concepts of the modeling technique the third detail level of the ETL-process modeling deviates from the original 3LGM² concept because it merges technical and concept layer. This is not inherent in the approach and could also be properly abstracted, but the proposed reference model merged it simply due to practical reasons.

As mentioned in the introductory section established tools supporting the GUI-based development and operation of ETL-processes exist. These were not found in the literature review although one could argue that such tools implement ETL-modeling techniques. However, ETL-process modeling techniques with a technical detail oriented perspective where not the focus of this work. The objective was to find a technique, which can help in management and maintenance of ETL-processes, as well as to serve as a kind of interface between regular IT-management, clinical data warehousing and secondary data users. Our findings indicate that the presented approach for modeling of ETL-processes and thereby processed information based on 3LGM² and openEHR can support that.

5. References

[1] Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, Peek N, Sacchi L, Secondary Use and Analysis of Big Data Collected for Patient Care, Yearb Med Inform. 2017 Aug;26(1):28-37, doi: 10.15265/IY-2017-008. Epub 2017 Sep 11. PubMed PMID: 28480474.

[2] Winter A, Brigl B, Wendt T, Modeling hospital information systems. Part 1: The revised three-layer graph-based meta model 3LGM2, Methods Inf Med. 2003;42(5):544-51, PubMed PMID: 14654889.

[3] Goossen W, Goossen-Baremans A, van der Zel M, Detailed clinical models: a review, Healthcare informatics research. 2010 Dec;16:201–214.

[4] Haarbrandt B, Gerbel S, Marschollek M, Einbindung von openEHR Archetypen in den ETL-Prozess eines klinischen Data Warehouse, GMDS 2014. 59. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS). Göttingen, 07.-10.09.2014. Düsseldorf: German Medical Science GMS Publishing House; 2014. DocAbstr. 230; 2014.

[5] Marco-Ruiz L, Moner D, Maldonado JA, Kolstrup N, Bellika JG, Archetype-based data warehouse environment to enable the reuse of electronic health record data. International journal of medical informatics, 2015 Sep;84:702–714.
[6] Kahn MG, Brown JS, Chun AT, Davidson BN, Meeker D, Ryan PB, et al., Transparent reporting of data quality in distributed data networks, EGEMS (Wash DC). 2015 Mar 23;3(1):1052. doi:10.13063/2327-9214.1052. eCollection 2015, PubMed PMID:25992385; PubMed Central PMCID: PMC4434997.
[7] Steiner J, Evaluation von Methoden zur Modellierung von ETL-Prozessen und der dabei verarbeiteten Daten am Beispiel eines großen Universitätsklinikums [master's thesis], TU Braunschweig, Braunschweig, 2017.
[8] HiGHmed | HiGHmed, http://www.highmed.org, last access: 29.01.2018.
[9] Trujillo J, Luján-Mora S, A UML based approach for modeling ETL processes in data warehouses, In: Song I-Y, Liddle SW, Ling T-W, Scheuermann P, editors. Conceptual Modeling, - ER 2003: Proceedings of the 22nd International Conference on Conceptual Modeling; 2003 Oct 13-16; Chicago, IL, USA. Berlin, Heidelberg:Springer; 2003. p. 307–20.
[10] Luján-Mora S, Trujillo J, A Comprehensive Method for Data Warehouse Design, In: Lenz H-J, Vassiliadis P, Jeusfeld MA, Staudt M, editors. Design and Management of Data Warehouses. DMDW 2003: Workshop Proceedings of the 5th International Workshop on Design and Management of Data Warehouses; 2003 Sep 8; Berlin, Germany. Aachen: CEUR-WS.org [contents]. p. 1.1-1.14.
[11] Muñoz L, Mazón J-N, Pardillo J, Trujillo J, Modelling ETL Processes of Data Warehouses with UML Activity Diagrams, In: Meersman R, Tari Z, Herrero P, editors. On the Move to Meaningful Internet Systems: OTM 2008 Workshops. Berlin: Springer; 2008. 44–53 p.
[12] Simitsis A, Modeling and managing ETL processes, In: VLDB PhD Workshop. Berlin; 2003.
[13] Dupor S, Jovanović V, An approach to conceptual modelling of ETL processes, In: Biljanovic , editor. Information and Communication Technology, Electronics and Microelectronics. MIPRO 2014: Proceedings of the 37th International Convention on Inf Commun Technol Electron Microelectron; 2014 May 26-30; Opatija, Croatia. New Jersey: IEEE; 2014. p. 1485–90.
[14] El Akkaoui Z, Zimanyi E, Defining ETL workflows using BPMN and BPEL, In: DOLAP '09 Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP. Hong Kong, China; 2009. 41–8 p.
[15] openEHR Foundation, openEHR – Clinical Models Program, http://www.openehr.org/programs/clinicalmodels/, last access: 13.12.2017.
[16] openEHR Foundation, Clinical Knowledge Manager, http://www.openehr.org/ckm/, last access: 07.03.2018.
[17] 3LGM²-Team, [3LGM²] – Startseite, http://www.3lgm2.de/, last access: 13.12.2017.
[18] Turley CB, Obeid J, Larsen R, Fryar KM, Lenert L, Bjorn A, Lyons G, Moskowitz J, Sanderson I, Leveraging a Statewide Clinical Data Warehouse to Expand Boundaries of the Learning Health System. EGEMS (Wash DC). 2016 Dec 6;4(1):1245. doi: 10.13063/2327-9214.1245. eCollection 2016. PubMed PMID: 28154834; PubMed Central PMCID: PMC5226381.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-212

Evaluation of SNOMED CT Content Coverage: A Systematic Literature Review

Farid KHORRAMI[a], Maryam AHMADI[a,1] and Abbas SHEIKHTAHERI[a]
[a] *Dept. of Health Information Management, School of health management and information sciences, Iran University of Medical Sciences, Tehran, Iran*

Abstract. Background: One of the most important features studied for adoption of terminologies is content coverage. The content coverage of SNOMED CT as a large scale terminology system has been evaluated in different domains by various methods. Objectives: This study provided an overview of studies evaluating SNOMED CT content coverage. Methods: This systematic literature review covered Scopus, Embase, PubMed and Web of Science. It included studies in English language with accessible full-text from the beginning of 2002 to November 2017. Results: Reviewing 62 studies revealed that 76 percent of studies were carried out in the US and other countries started to study in this regard from 2007. Most of the studies focused on the comparison of SNOMED CT with disease classifications in the domain of "diagnosis and problem list". Conclusion: Studying the trend of studies in different countries shows that SNOMED CT content coverage is not limited to the early stages of SNOMED CT adoption. However, evaluation methods are likely different due to the stage of SNOMED CT implementation. Therefore, it is recommended to identify and compare evaluation methods of SNOMED CT content coverage in future studies.

Keywords: Systematized Nomenclature of Medicine, Terminology, Terminological systems, Content coverage, Evaluation

1. Introduction

Applying terminological systems (TS) at the time of data capture is one of the requirements for documentation in electronic information systems [1,2]. Terminologies have a wide range of applications such as user interface for knowledge resources including guidelines [3], critical pathways and reminders; decision support systems[4] and practice analysis support in quality improvement, epidemiology and outcomes analyses [1,5-7]. Capturing data based on TS allows for data reuse [5,6,8]. Hence, it is critical to 1) select and implement suitable TS and 2) evaluate it after adoption.

SNOMED CT has been used as both reference terminology [9-11] and interface terminology [12,13] in information systems as well as research studies. This terminology is known as the primary infrastructure for data interchange in EHR. SNOMED CT usages in different studies are classified into 15 categories and five main focus categories [14]. One of the focus categories is "evaluatio”" which is divided into two subgroups: 1) "Prove merit" dealing with studies evaluating the advantages of SNOMED CT in operational settings, and 2) "Retrieval or analysis of patient data" which deals with

[1] Corresponding Author: Maryam Ahmadi, School of Health Management and Information Sciences, Iran University of Medical Sciences, Address: No. 6, Rashid Yasemi St., Vali-e-asr Ave., 1995614111 Tehran, Iran. Tel.: +98 9171982334; fax: +98 21 88883334 , E-Mail: mahmadi24@gmail.com

studies evaluating SNOMED CT used for capturing data in patient care. In the latter case, the approach is changing from a focus on using SNOMED CT for capturing data to using "the captured data"[14]. As ASSESS CT project of EU studied the fitness of SNOMED CT to be used as a potential standard for EU-wide eHealth deployments [15]. One of the most important issues in studies of this subgroup is "content coverage" [7,16]. In fact, it is the "quintessential feature" of TS [17]. Having had a wide range of applications and considering its development in more than thirty countries [18], SNOMED CT evaluation methods for content coverage need to be identified and evaluated to assure its development in parallel with the growing implementation of TS in health care information systems. In a bigger study, we have focused on this matter. So as part of it, in this paper, we are going to provide an overview of studies evaluating SNOMED CT content coverage.

2. Methods

2.1. Search Strategy

Databases including PubMed, Scopus, Embase and Web of Science were searched without time limitation. The Mesh term "Systematized Nomenclature of Medicine" and all its entry terms were searched with “OR” operator. In addition, we searched different terms for evaluation such as Evaluat*, Assess*, analys*, analyz*, test* and audit* [19] using "OR" operator. Additionally, the term "coverage" was also searched. All of the three search strategies were conducted in titles, abstracts and keywords. Finally, the results of the above searches were combined together using "AND" operator. The research team also checked the references of the retrieved articles to find any related articles missed through the searching process.

2.2. Study selection and eligible papers

We included English papers dealing with content coverage evaluation in SNOMED CT from 2002 (the first edition of SNOMED CT) to 2017. Papers whose full-text was not accessible and papers related to previous version of SNOMED were excluded.

Study selection steps were as per PRISMA flow diagram. After importing the selected articles into the EndNote, the duplicate items were excluded. Two members of the research team (A.ST and F.K) separately screened all the titles and the abstracts of the papers. Papers were separately grouped as 1) relevant, 2) need to more review, and 3) not relevant. The two researchers discussed over the cases of disagreement in presence of the third member (M.A) of the team. In the next step, the two researchers studied the papers in groups 1 and 2 independently. Having reached consensus, the research team selected related papers meeting the objectives of the study.

2.3. Data extraction and summarizing

Required data were extracted from the selected articles based on data elements by the two researchers (A.ST and F.K). For disagreement cases, the two researchers discussed over items extracted from the articles in presence of the third researcher (M.A). Finally, they agreed on the cases based on majority vote.

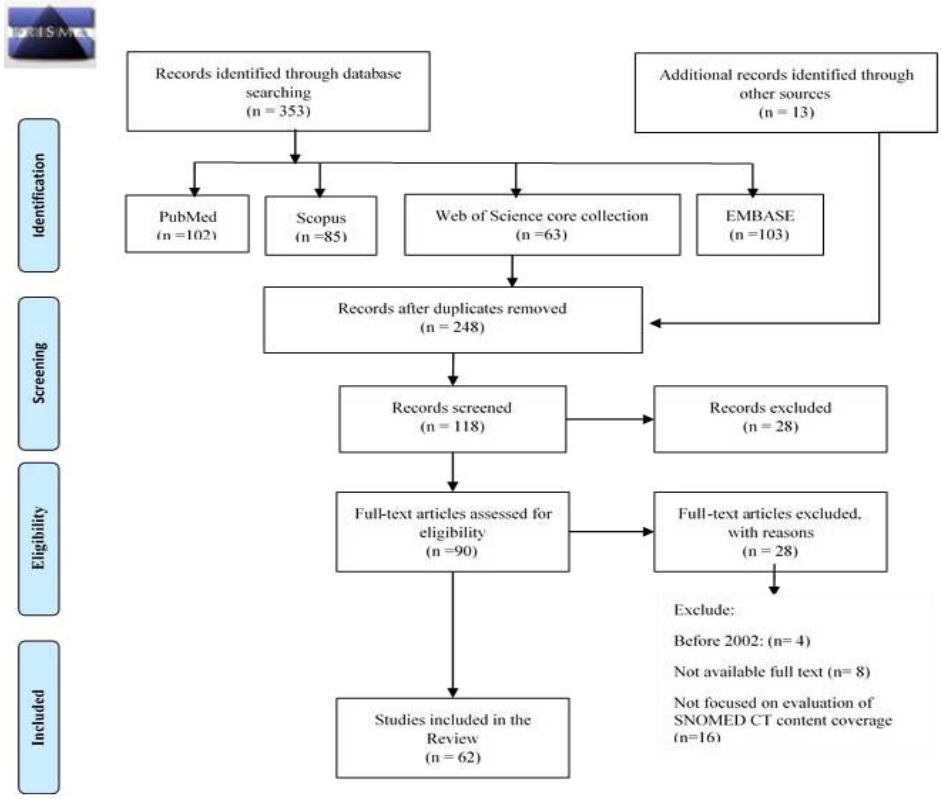


Figure 1. Details of selection of the studies based on PRISMA flow diagram.

3. Results

Out of 366 papers retrieved through the search strategy and studying the resources, we obtained 118 papers excluding duplicated items. After screening the titles and abstracts, 28 more papers were excluded from the study. Therefore, 90 papers remained for full-text study. Twenty-eight papers were also excluded from the study: 26 papers after applying the exclusion criteria and 2 papers through data extraction. Ultimately, 62 publications were selected for the final analysis. Figure 1 depicts the details of selection of the studies based on PRISMA flow diagram.

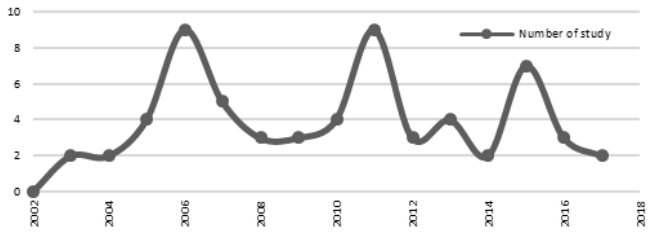


Figure 2-Trend of studies from 2002 to 2017

Table 1. The frequency and percentage of published papers by country

Country	n	%
USA	47	76%
Netherlands	4	6%
Sweden	3	5%
Korea	2	3%
Canada	1	2%
France	1	2%
Germany	1	2%
Denmark	1	2%
Argentina	1	2%
Kenya	1	2%
Total	**62**	**100%**

Figure 2 shows the trend of studies from 2002 to 2017. Most of studies were carried out in 2006 and 2011.Twenty-four percent of the selected studies were presented in conferences. AMIA Annual Symposium Proceedings and Journal of the American Medical Informatics Association had the maximum share of published articles (45 percent).

3.1. Countries

Most of the studies were carried out in high income countries. The extracted data showed that 76 percent were published in the US and 16 percent in Europe. However, the Netherlands had the highest share (50%) in Europe (Table 1).

Having in mind the first edition of SNOMED CT published in 2002 and its internationalization in 2007, Table 2 shows the trend of the studies in different years by country in three periods. Seventy-three percent of the total studies were carried out after internationalization of the system (2007), and this figure was 86 percent in other countries apart from the US.

3.2. Domain

Table 3 shows the results by the evaluated domains. The domain with maximum shares included Diagnosis and Problem List and Radiology. The focus of all studies out of the US was Diagnosis and Problem List, Nursing and Oncology.

3.3. Terminological systems

Studies show that the evaluation of SNOMED CT content coverage was carried out either by itself or in comparison with one or more terminology systems. The findings of this research are:

- Evaluating the content coverage of SNOMED CT by itself: 29% (18 cases)
- Evaluating the content coverage of SNOMED CT with another system: 34% (21 cases)
- Comparing the content coverage of several terminology systems including SNOMED CT: 37% (23 cases)

Table 2. The frequency of papers published in three different time periods by grouping of countries

Years range	Total		USA		Other countries	
	n	%	n	%	n	%
2002-2006	17	27%	15	32%	2	13%
2007-2011	24	39%	16	34%	8	53%
>2012	21	34%	16	34%	5	33%
Total	**62**		**47**		**15**	

Table 3. The frequency of domain evaluation in SNOMED CT content coverage

Domains groups	n	%
Diagnosis and Problem list	34	51%
Nursing	4	6%
Radiology	4	6%
Ophthalmology	2	3%
Medical Guidelines	2	3%
Adverse drug reactions	1	2%
Dental finding	1	2%
Oncology	4	6%
Clinical research concepts	4	6%
Allergy	2	3%
Disability	2	3%
Herbal terms	1	2%
Phonemics	1	2%
Total	**62**	

Table 4 shows the terminology systems had been studied. Most of evaluation studies were carried out against disease classification, UMLS and LOINC with 26, 24, and 10 percent, respectively.

Fourteen studies also addressed the comparison of SNOMED CT with interface terminologies which; in fact, was a combination of various terminology systems based on the user needs.

Thirty-nine percent of papers did not mention the version of SNOMED CT or UMLS clearly.

Table 4. The frequency of other terminology systems in SNOMED CT content coverage

Category title (total)	
Classification of diseases (16)	**n**
ICD-9-CM	7
ICD-10-CM	2
ICD-10	4
ICD-11	1
ICD-O-3	1
ICF	1
UMLS (12)	
Medication and Allergy terminologies (9)	
MedDRA	3
RxNorm	2
NDF-RT	2
UNII	2
LOINC (6)	
Imaging terminologies (4)	
RadLex	4
Interface terminologies(14)	
Medcin	4
Veterans Benefit Administration codes (VBA)	3
Cancer Data Standards Registry and Repository (CaDSR)	1
Categorical Health Information Structured Lexicon (CHISL)	1
Healthcare Data Dictionary (HDD)	1
Phenoslim terminology (PT)	1
Mayo clinic vocabulary	1
Diagnoses for Intensive Care Evaluation (DICE)	1
Nursing terminologies (3)	
ICNP	3
Other terminologies (4)	
CPT	2
MESH	2

4. Discussion

The aim of this review is to provide a general overview of articles addressing evaluation of SNOMED CT content coverage. The obtained results can generally be classified into four main categories: 1) Place of the studies (countries), 2) Trend of studies in the time limit of this study, 3) Terminology systems compared with SNOMED CT in studies of content coverage evaluation and 4) Domains evaluated.

Studies addressing evaluation of SNOMED CT content coverage were carried out in a few countries with high income. Seventy-six percent of the studies were in the US. The contribution of this country before 2007 (before internalization of SNOMED CT) was 88%. However, in an earlier review article, Lee et al (2014) who studied the usage of SNOMED CT from 2001 to 2012 showed that the contribution of the US was 53%[14]. Our study showed that the contribution of European countries was 16%. Fifty percent of the studies in Europe were published by the Netherlands. The significant point was that there were no studies from England and Australia in this subject. However, Lee et al (2014) reported the two countries along with France had maximum number of articles after the US [14]. The countries did not have any studies in the evaluation domain.

Studying the trend of studies carried out within the period of our study shows that content coverage evaluation has grown since 2007. Ninety-three percent of the total studies in all countries excluding the US were carried out after 2007. Since content coverage studies are included in SNOMED CT predevelopment [14], it is natural for new country members of IHTSDO to study on this domain at initial years of membership. For example, combination of some studies [3; 18-24] in the Netherlands led to the development of an evaluation framework for characterizing terminological systems[17]. It is a good example of specialized follow-up for SNOMED CT adoption in this country.

In some of studies evaluating SNOMED CT content coverage, other TS were also used. Studying compared to SNOMED CT showed that they were either carried out alone or in comparison with other TS. They include SNOMED CT content coverage evaluation compared to medical record texts or clinical cases (29%), and content coverage compared to other TS (81%). The systems for diseases classification had the highest portion (26%). It can be the weakness of classification systems in covering specific terms at the time of clinical data capturing [25]. Moreover, most of studies in this topic introduce SNOMED CT terminology systems as the best terminology system for improving content coverage in different domains [9; 13; 26-29]. Some of the studies compared SNOMED CT system with interface terminologies such as Medcin [12; 13; 30], CHISL [13] and Mayo clinic [16; 31]. Since such TS are used for systematic recording of data in most IHTSDO member and non-member countries, development of a framework for their evaluation is necessary. This issue is technically discussed in [32]. UMLS – as a terminological system and also a controlled meta-thesaurus – is known for creating interoperability between computerized systems [33]. Twenty-four percent of studies directly mentioned that they had used UMLS tools in different stages of their study. It can, therefore, be implied that UMLS tools not only affect on interoperability but also on terminological systems.

SNOMED CT covers a wide range of domains. The findings of this research shows that most of the studies were carried out in "Diagnosis and problem list"(51%). This confirms Studying [14] the evaluated domains shows that selecting them gets more specialized through the years and the studies focus on a disease advanced findings. For example, studies [34-36] concentrate on content coverage of a specific disease.

We just reviewed articles in English with accessible full-text. However, there might been studies addressing content coverage of SNOMED CT in other languages which we missed.

5. Conclusion

Our review showed that the number of studies addressing evaluation of SNOMED CT content coverage is growing and the topic is considered by different countries which most of them are high income countries. Although content coverage is basically carried out at the preliminary stages of SNOMED CT adoption, our study revealed that such studies do not stop; however, the evaluation methods may change. Therefore, the identification and comparison of SNOMED CT content coverage – which is being carried out as a future study.

Acknowledgements

This study is a part of a PhD dissertation granted by Iran University of Medical Sciences (Grant No: IUMS/SHMIS_2017/9321563002).

6. References

[1] D. Bronnum, A.R. Hojen, K.R. Goeg, and P.B. Elberg, Terminology-Based Recording of Clinical Data for Multiple Purposes Within Oncology, Stud Health Technol Inform 228 (2016), 267-271.
[2] M. Dugas, F. Fritz, R. Krumm, and B. Breil, Automated UMLS-Based Comparison of Medical Forms, Plos One 8 (2013).
[3] L. Ahmadian, N.F. De Keizer, and R. Cornet, The use of SNOMED CT for representing concepts used in preoperative guidelines, Stud Health Technol Inform 150 (2009), 658-662.
[4] Y. Lin, C.J. Staes, D.E. Shields, V. Kandula, B.M. Welch, and K. Kawamoto, Design, Development, and Initial Evaluation of a Terminology for Clinical Decision Support and Electronic Clinical Quality Measurement, AMIA Annu Symp Proc 2015 (2015), 843-851.
[5] C.M. Cusack, G. Hripcsak, M. Bloomrosen, S.T. Rosenbloom, C.A. Weaver, A. Wright, D.K. Vawdrey, J. Walker, and L. Mamykina, The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting, J Am Med Inform Assoc 20 (2013), 134-140.
[6] A. Randorff Hojen and K. Rosenbeck Goeg, Snomed CT implementation. Mapping guidelines facilitating reuse of data, Methods Inf Med 51 (2012), 529-538.
[7] S.H. Brown, P.L. Elkin, E. Lumakovska, M. Rallins, T. Rosenbloom, J. Trajkovski, and M.S. Tuttle, Terminology and Terminological Systems, Springer, London, 2012.
[8] A. Randorff Hojen, E. Sundvall, and K. Rosenbeck Goeg, Visualizing sets of SNOMED CT concepts to support consistent terminology implementation and reuse of clinical data, Stud Health Technol Inform 192 (2013), 1160.
[9] S.H. Brown, C.S. Husser, D. Wahner-Roedler, S. Bailey, L. Nugent, K. Porter, B.A. Bauer, and P.L. Elkin, Using SNOMED CT as a reference terminology to cross map two highly pre-coordinated classification systems, Stud Health Technol Inform 129 (2007), 636-639.
[10] K.A. Spackman, Integrating sources for a clinical reference terminology: experience linking SNOMED to LOINC and drug vocabularies, Medinfo 9 Pt 1 (1998), 600-603.
[11] S. Schulz, J.M. Rodrigues, A. Rector, and C.G. Chute, Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration, Stud Health Technol Inform 245 (2017), 940-944.
[12] S.H. Brown, S.T. Rosenbloom, B.A. Bauer, D. Wahner-Roedler, D.A. Froehling, K.R. Bailey, M.J. Lincoln, D. Montella, E.M. Fielstein, and P.L. Elkin, Direct comparison of MEDCIN and SNOMED CT for representation of a general medical evaluation template, AMIA Annu Symp Proc 11 (2007), 75-79.

[13] S.T. Rosenbloom, S.H. Brown, D. Froehling, B.A. Bauer, D.L. Wahner-Roedler, W.M. Gregg, and P.L. Elkin, Using SNOMED CT to represent two interface terminologies, J Am Med Inform Assoc 16 (2009), 81-88.
[14] D. Lee, N. de Keizer, F. Lau, and R. Cornet, Literature review of SNOMED CT use, J Am Med Inform Assoc 21 (2014), e11-19.
[15] ASSESS CT - Assessing SNOMED CT for Large Scale eHealth Deployments in the EU, in, 2018.
[16] P.L. Elkin, S.H. Brown, A. Balas, Z. Temesgen, D. Wahner-Roedler, D. Froehling, M. Liebow, B. Trusko, S.T. Rosenbloom, and G. Poland, Biosurveillance evaluation of SNOMED CTś terminology (BEST Trial): Coverage of chief complaints, in: Stud Health Technol Inform, 2008, pp. 797-802.
[17] R. Cornet, N.F. de Keizer, and A. Abu-Hanna, A framework for characterizing terminological systems, Methods Inf Med 45 (2006), 253-266.
[18] R. Cornet and N. de Keizer, Forty years of SNOMED: a literature review, BMC Med Inform Decis Mak 8 Suppl 1 (2008), S2-S2.
[19] D.G.T. Arts, R. Cornet, E. De Jonge, and N.F. De Keizer, Methods for evaluation of medical terminological systems: A literature review and a case study, Methods Inf Med 44 (2005), 616-625.
[20] R. Cornet, Do SNOMED CT relationships qualify?, Stud Health Technol Inform 136 (2008), 785-790.
[21] L. Ahmadian, R. Cornet, W.A. van Klei, and N.F. de Keizer, Diversity in preoperative-assessment data collection, a literature review, Stud Health Technol Inform 136 (2008), 127-132.
[22] F. Bakhshi-Raiez, R. Cornet, and N.F. de Keizer, Development and Application of a Framework for Maintenance of Medical Terminological Systems, J Am Med Inform Assoc 15 (2008), 687-700.
[23] R. Cornet and S. Schulz, Relationship groups in SNOMED CT, Stud Health Technol Inform 150 (2009), 223-227.
[24] R. Cornet, Definitions and qualifiers in SNOMED CT, Methods Inf Med 48 (2009), 178-183.
[25] K.W. Fung and J. Xu, An exploration of the properties of the CORE problem list subset and how it facilitates the implementation of SNOMED CT, J Am Med Inform Assoc 22 (2015), 649-658.
[26] H. Wasserman and J. Wang, An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list, AMIA Annu Symp Proc (2003), 699-703.
[27] T. Sampalli, M. Shepherd, J. Duffy, and R. Fox, An evaluation of SNOMED CT (R) in the domain of complex chronic conditions, Int J Integr Care 10 (2010).
[28] P.L. Elkin, S.H. Brown, A. Balas, Z. Temesgen, D. Wahner-Roedler, D. Froehling, M. Liebow, B. Trusko, S.T. Rosenbloom, and G. Poland, Biosurveillance evaluation of SNOMED CT's terminology (BEST Trial): coverage of chief complaints, Stud Health Technol Inform 136 (2008), 797-802.
[29] D.A. Travers, S.W. Haas, D.A. Travers, and S.W. Haas, Unified Medical Language System Coverage of Emergency-medicine Chief Complaints, Acad Emerg Med 13 (2006), 1319-1323.
[30] D. Montella, S.H. Brown, P.L. Elkin, J.C. Jackson, S.T. Rosenbloom, D. Wahner-Roedler, G. Welsh, B. Cotton, O.D. Guillamondegui, H. Lew, K.H. Taber, L.A. Tupler, R. Vanderploeg, and T. Speroff, Comparison of SNOMED CT versus Medcin terminology concept coverage for mild Traumatic Brain Injury, AMIA Annu Symp Proc 2011 (2011), 969-978.
[31] P.L. Elkin, S.H. Brown, B.A. Bauer, C.S. Husser, W. Carruth, L.R. Bergstrom, and D.L. Wahner-Roedler, A controlled trial of automated classification of negation from clinical notes, BMC Med Inform Decis Mak 5 (2005).
[32] S.T. Rosenbloom, R.A. Miller, K.B. Johnson, P.L. Elkin, and S.H. Brown, A model for evaluating interface terminologies, J Am Med Inform Assoc 15 (2008), 65-76.
[33] S.T. Wu, H. Liu, D. Li, C. Tao, M.A. Musen, C.G. Chute, and N.H. Shah, Unified Medical Language System term occurrences in clinical notes: a large-scale corpus analysis, J Am Med Inform Assoc 19 (2012), e149-156.
[34] D. Brammen, H. Dewenter, K.U. Heitmann, V. Thiemann, R.W. Majeed, F. Walcher, R. Rohrig, and S. Thun, Mapping Equivalence of German Emergency Department Medical Record Concepts with SNOMED CT After Implementation with HL7 CDA, Stud Health Technol Inform 243 (2017), 175-179.
[35] M.S. Park and Z. He, Consumers' Use of UMLS Concepts on Social Media: Diabetes-Related Textual Data Analysis in Blog and Social Q&A Sites, JMIR Med Inform 4 (2016), e41.
[36] T. Oluoch, N. de Keizer, P. Langat, I. Alaska, K. Ochieng, N. Okeyo, D. Kwaro, and R. Cornet, A structured approach to recording AIDS-defining illnesses in Kenya: A SNOMED CT based solution, J Biomed Inform 56 (2015), 387-394.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-220

Nurses' Attitude for Using Barcode Medication Administration System in a Developing Country

Abbas SHEIKHTAHERI[a,b] and Samane SARAVANI-AVAL[c,1]
[a] *Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran*
[b] *Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran*
[c] *Lecturer of Health Information Technology Dept., School of Allied Medical Sciences, Zabol University of Medical Sciences, Zabol, Iran*

Abstract. Medication errors are common in healthcare settings. To prevent these errors, use of modern technology is suggested. Improvement of medication administration system particularly at the time of drug administration is mandated in Iran. Barcode medication administration (BCMA) systems are useful in this regard. This study was conducted to assess nurses' attitude for the use of BCMA systems. To this end, 283 randomly selected nurses working in teaching hospitals were surveyed using a five point Likert scale questionnaire. The most positive attitudes were related to the role of BCMA in reducing medication errors (4.02±0.8), improving performance (3.8±0.99), productivity (3.8±0.97) and making patient care easier (3.83±0.9). Only 9.9 percent of nurses did not like to use the technology. There was no significant difference between nurses in terms of their age, experience, education, and hospitals. In conclusion, **n**urses' attitude about BCMA was in a relatively good level. However, there is a need to train nurses for the use of this technology.

Keywords. Barcode medication administration system, Medication errors, Medication systems, Nurses, Readiness assessment

1. Introduction

Patient safety is one of the most important aspects of healthcare quality [1]. One of the critical aspects of healthcare services is the medication administration process. Medication errors are among the most frequent and important patient safety threats. Medication errors are the eighth leading cause of death in America [2] and may occur in each of the medication process steps including prescription medication, transcription of physician's orders, drug dispensing, administering the drugs to patients or controlling the drug effects [3]. A study in Iran showed that nurses spend more than 40 percent of their working time on giving drugs to patients, so maintaining safety and preventing medication errors in the medication administration process is of special importance [4]. A study has shown that medication errors for every nurse have been 19.5 cases per nurse on average within three months [5].

[1] Corresponding author: Samane Saravani-Aval. School of Allied Medical Sciences, Zabol University of Medical Sciences, Zabol, Iran. Tel:(+98) 54-32232166. E-mail: s.saravani.aval@gmail.com

The use of barcode medication administration system (BCMA) is suggested for improving patient safety in terms of medication errors [6, 7]. BCMA controls "five rights" related to medication administration, including the right patient, the right drug, the right time, the right route and the right dose [8]. One study about this technology have shown improvement in medication errors caused by incorrect drug prescription (75%), patient misidentification (93%) and drug misidentification (62%) [9]. Other studies have also shown the effectiveness of this technology for reducing medication errors [10].

The US Food and Drug Administration obliged the use of medication bar-coded labels and predicted that many medication errors will be prevented with this obligation [11]. According to the "Iranian Evaluation Guide of Clinical Governance" published by the Ministry of Health, hospitals should identify patients with at least two identifications. Another standard in this regard, is establishment of a medication administration system particularly at the time of drug administration [12]. Therefore, BCMA is a good technology for these standards.

Despite the benefits, adoption and acceptance of users are important. From previous studies, it was found that one of the main causes of system failure is the lack of participation and readiness of users [13]. Nurses have an important role in implementing health information technologies [14], especially BCMA, because they are the main users of BCMA. No studies have addressed the attitude of nurses to implement and use of this technology in Iran, therefore, this study was conducted with the aim of assessing the nurses' attitude in teaching hospitals in Iran.

2. Method

Three hundred nurses from seven teaching hospitals affiliated with Iran University of Medical Sciences (IUMS), Tehran, Iran were invited randomly in 2016 to participate in the study. The sample size for each hospital was determined based on the number of nurses in each hospital. Finally 283 nurses (94%) participated.

We used a questionnaire with eight questions in the form of five-point Likert scale. This study was conducted before implementation of BCMA technology to explore nurse attitude for future use. Therefore, the part 1 of the questionnaire was devoted to introduction of BCMA. In this part, we described the technology and related workflows to enable nurses to perceive the technology and its effects on their workflows. Attitude questions were designed based on several different studies [15, 16]. Content validity of the questionnaire was approved by 11 faculty members of the fields of nursing, health information management, medical informatics and a number of nurses working in hospitals (out of the sample). Reliability was assessed using Cronbach's alpha coefficient, so that 30 nurses out of the sample completed the questionnaire before the main study ($\alpha = 0.752$). We distributed the questionnaires among the nurses and collected the completed ones several days after distribution. The Ethics Committee of IUMS approved this study.

We scored the responses as follows: 1 for completely disagreed to 5 for completely agreed. Negative questions were scored inversely. The data were analyzed using descriptive (frequency, percentage and mean) and inferential statistics (non-parametric tests after checking the normality of data using Kolmogorov-Smirnov test) using SPSS 20. We classified score as weak (<25% of score), moderate (25-50% of score), relatively good (50-75% of score) and good (>75% of score).

Table 1. Nurses' attitude towards using BCMA

Statements	Completely agreed	Agreed	So- So	disagreed	Completely disagreed	Mean± SD
It improves job performance.	67(23.7)	143(50.5)	35(12.4)	30(10.6)	7(2.5)	3.8±0.99
It increases my productivity.	66(23.3)	138(48.8)	45(15.9)	28(9.9)	6(2.1)	3.8±0.97
It enhances effectiveness in job.	61(21.6)	141(49.8)	47(16.6)	29(10.2)	5(1.8)	3.79±0.95
It is fast to use in my job.	53(18.7)	153(54.1)	34(12)	36(12.7)	6(2.1)	3.75±0.97
It reduces medication errors.	78(27.6)	158(55.8)	26(9.2)	18(6.4)	3(1.1)	4.02±0.84
It makes caring for patients easier.	59(20.8)	152(53.7)	39(13.8)	29(10.2)	3(1.1)	3.83±0.91
I do not want BCMA change the way I work.	27(9.5)	61(21.6)	69(24.4)	101(35.7)	25(8.8)	3.1±1.1
I would like to use this system.	77(27.2)	140(49.5)	34(12)	20(7.1)	8(2.8)	3.92±0.97

3. Results

Among the participants, 85.9% of nurses were female. The mean and standard deviation (SD) of nurses' age were 32.3±6.5 years and their average working experience was 7.7±5.9 years. Most of them had a bachelor degree (94.3%) and 147 of them participated from general hospitals. Among them, 75.6% had participated in patient safety courses and 36% of them had courses or workshops about application of IT for patient safety.

According to Table 1, the most positive attitude of nurses was about reducing medication errors (4.02 ± 0.8), improving job performance (3.8 ± 0.9), improving productivity (3.8 ± 0.97) and making patient care easier (3.83 ± 0.9). 9.9 percent of the nurses did not like to use BCMA technology. The lowest score was for nurses' reluctance to changing their workflow using the technology. The average total score (30 ± 5.9 from 40) showed that overall attitude of nurses was relatively good (75% of possible score). According to the table 2, nurses in general and specialized hospitals have similar attitude. The results also showed that the most positive attitude in general and specialized hospitals was related to the reduction of medication errors using BCMA and tendency to use the system. Comparing between the two hospital groups showed that there was only a significant difference between nurses' attitude towards the speed of performing tasks using the system (p=0.007).

Table 2. Nurses' attitude in general and specialized hospitals towards using BCMA

Statements	General	Specialized	P-value
Bar coding system improves job performance.	3.9 ±0.9	3.78±1	0.415
Bar coding system increases productivity.	3.88±0.9	3.74±1	0.279
Bar coding system enhances effectiveness in job.	3.83±0.9	3.75±0.9	0.434
Bar coding system is fast to use in my job.	3.9 ±0.8	3.59±1	0.009
Bar coding system reduces medication errors.	4.05±0.8	4 ±0.9	0.804
Bar coding system makes caring for patients easier.	3.86±0.84	3.8 ±0.9	0.777
Overall, I don't want the BCMA change the way I currently work.	3.14±1.1	3.11±1.2	0.755
I would like to use this system.	3.98±0.9	3.87±0.9	0.212
Total score	**30.4±5.7**	**29.6±6.2**	**0.144**

4. Discussion

Ackerman et al argue that understanding human factors and its relations to technology and in fact, the socio-technical nature of health information technologies is a prerequisite of successful implementation of a technology in health settings [17]. Based on the studies, users' perception is an influential factor for technology acceptance [18]. Other researchers introduce technology support and end users' perception as the effective factors for BCMA implementation [19]. In this regard, we found that nurses' attitudes were at a relatively good level and also there were no statistical differences between nurses' attitude in general and specialized hospitals and their demographic and experience did not affect on their attitude.

A study showed that two effective factors in BCMA acceptance by users are "understanding the profitability" and "ease of use" [20]. The present study showed that nurses have a good understanding about the usefulness of BCMA and believed that BCMA is suitable to reduce errors and improve performance and effectiveness. Studies showed that improvement is observed in medication errors using BCMA [9, 10]. Another study found that application of information technology increases nurses' productivity and using the technology may save nurses' time for patient care [21]. Our nurses have such an attitude about their productivity using BCMA. Other studies in other countries showed that nurses have a good attitude towards using BCMA [22,15]. These studies have similar findings to our results.

Minda found that the time required for documentation by computers is considerably less than the time needed for manual documentation [23]. Our study showed that nurses have a positive attitude but they are concerned about the speed of the technology, because some nurses believed that the steps of nursing and patient care processes are increased after using BCMA. Holden showed that nurses need a lot of support and training to accept this technology [15], we observed that in the majority of hospitals there were no suitable training courses for this technology. Hostgaard showed that the use of specific training will accelerate the pace of acceptance and will finally create a positive outlook in nurses [24]. Mary et al. have emphasized on nurses' training about BCMA using different modes such as guidelines, and tutorials [25]. Previous Iranian studies showed that nurses considered training about the technology very important [14]. In addition, other Iranian studies have evaluated the level of computer skills of nurses, which indicates the appropriate skills of nurses [26]. Therefore, we believe that the nurses need to be more trained about the BCMA and its processes to better understand the effect of this technology on their workflows in order to decrease this concern about the speed of the technology.

We have some limitations that should be considered. First, Iranian studies are in early steps of implementing patient safety programs and using technologies for patient safety. The barcode technology has not been implemented in this country for medication administration; however, barcodes are currently used for other processes such as admission. Therefore, the findings may not reflect the real attitude of nurses. However, we believe that investigating nurses' attitude before implementing a new technology is also important to predict future use of the technology. Additionally, although our questions were based on technology acceptance theories, we did not apply these theories to investigate effective factors on nurses' attitude. Applying such theories provide more useful information for implementing the BCMA. In a bigger follow-up study, we are currently conducting such investigation. In addition, although we have a large sample

size, they were from seven teaching hospitals. Therefore, our findings may not be generalizable to the country.

In conclusion, nurses play a key role in the successful implementation of BCMA and we concluded that nurses have a positive view of this technology and tend to use it. Therefore, hospitals in Iran may develop plan to implement and use the BCMA technology, because this technology seems acceptable for nurses. However, implementing this technology requires funds, training and also organizational readiness. In addition, other main users such as pharmacists' perspectives should be considered. Therefore, other studies should be conducted to assess the organizational readiness and identify the weaknesses in this regard. Additionally, training nurses regarding BCMA is also necessary.

Acknowledgments

This study is a part of a research financially supported by Iran University of Medical Sciences (Code: 94-04-136-26898).

References

[1] A. Sheikhtaheri, F. Sadoughi, M. Ahmadi, A framework of a patient safety information system for Iranian hospitals: lessons learned from Australia, England and the US, *International Journal of Medical Informatics* **82** (2013), 335-344.

[2] F. Ebrahimpour, A. Shahrokhi, A. Ghodousi, Patients' safety and nurses' medication administration errors, *Journal of Forensic Medicine* **20** (2014), 401-408. [In Persian]

[3] F. Valizadeh, S.F. Ghasemi, S.S. Nagafi, B. Delfan, A. Mohsenzadeh A Errors in medication orders and the nursing staff's reports in medical notes of children, *Iranian Journal of Pediatrics* **18** (2008) (Suppl 1), 33-40.

[4] S. Shahri, A. Kebriaee, H. Seyedi, Z. Sarafraz, Patient safety climate in medical centers of Kashan, *Journal of Health Promotion Management* **1** (2012), 62-72.

[5] M. Mirzaei, A. Khatony, R.S. Faramani, E. Sepahvand, Prevalence, types of medication errors and barriers to reporting errors by nurses in an educational hospital in Kermanshah, *Hayat* **19** (2013), 28-37.

[6] L. Harrington, K. Clyne, M.A. Fuchs, V. Hardison, C. Johnson C, Evaluation of the use of bar-code medication administration in nursing practice using an evidence-based checklist, *Journal of Nursing Administration* **43** (2013), 611-617.

[7] J. Bubalo, B.A. Warden, J.J. Wiegel, T. Nishida, E. Handel, L.M. Svoboda, L. Nguyen, P.N. Edillo, Does applying technology throughout the medication use process improve patient safety with antineoplastics? *Journal of Oncology Pharmacy Practice* **20** (2014), 445-460.

[8] G.L. Cochran, K.J. Jones, J. Brockman, A. Skinner, R.W. Hicks Errors prevented by and associated with bar-code medication administration systems, *Joint Commission Journal on Quality and Patient Safety* **33** (2007), 293-301.

[9] A.A. Leung, C.R. Denham, T.K. Gandhi, A. Bane, W.W. Churchill, D.W Bates, E.G. Poon, A safe practice standard for barcode technology, *Journal of Patient Safety* **11** (2015), 89-99.

[10] J. Bonkowski, C. Carnes, J. Melucci, J. Mirtallo, B. Prier, E. Reichert, S. Moffatt Bruce, R. Weber, Effect of barcode assisted medication administration on emergency department medication errors, *Academic Emergency Medicine* **20** (2013), 801-806.

[11] Patterson ES, Rogers ML, Render ML (2004) Fifteen best practice recommendations for bar-code medication administration in the Veterans Health Administration. The Joint Commission Journal on Quality and Patient Safety 30 (7):355-365

[12] S. Emami Razavi, H. Ravaghi, M. Mohaghegh, M. Sadat, F. Mostofian, S. Vazirian, *Assessment of patient safety in hospitals: a manual for evaluators,* Tehran: Mehr Ravesh, 2012.[In Persian]

[13] N.M. Lorenzi, R.T. Riley, Organizational aspects of health informatics: managing technological change, Springer Science & Business Media , 2013.

[14] K. Kimiafar, A. Sheikhtaheri, M. Sarbaz, Prioritizing factors influencing nurses' satisfaction with hospital information systems: a fuzzy analytic hierarchy process approach, *Comput Inform Nurs* **32** (2014), 174-181.
[15] R.J. Holden, R.L. Brown, M.C. Scanlon, B.T Karsh, Modeling nurses' acceptance of bar coded medication administration technology at a pediatric hospital. *Journal of the American Medical Informatics Association* **19** (2012), :1050-1058.
[16] M.M. VanderKooi, An evidence-based evaluation of medication barcode scanning acceptance in a community hospital, 2014. Available from https://scholarworks.gvsu.edu/dissertations/22/ Accessed 29/01/2018.
[17] S.L. Ackerman, K. Tebb, J.C. Stein, B.W. Frazee, G.W. Hendey, L.A. Schmidt, R.Gonzales, Benefit or burden? A sociotechnical analysis of diagnostic computer kiosks in four California hospital emergency departments. *Social Science & Medicine* **75** (2012), 2378-2385.
[18] S. Ajami, S. Ketabi, S.S. Isfahani, A. Heidari, Readiness assessment of electronic health records implementation. Acta Informatica Medica **19** (2011), 224-227.
[19] B. Lesselroth, J. Yang, J. McConnachie, T. Brenk, L. Winterbottom, Addressing the sociotechnical drivers of quality improvement: a case study of post-operative DVT prophylaxis computerized decision support. *BMJ Quality & Safety* **20** (2011), 381-389 .
[20] V. Taliercio, B. Schachner, D. Borbolla, D. Luna, E. Villalba, The expectations of nurses about the implementation of a barcoded medication administration system: a qualitative study, *Stud Health Technol Inform* **205** (2014), 191-195.
[21] L. Banner, C.A. Oleny, A step in the right direction: electronic clinical documentation improves nurse charting, efficiency, and satisfaction, *Comput Inform Nurs,* **25** (2007), 307-312.
[22] N. Samaranayake, S. Cheung, K. Cheng, K. Lai, W. Chui, B. Cheung, Implementing a bar-code assisted medication administration system: effects on the dispensing process and user perceptions. *International Journal of Medical Informatics* **83** (2014), 450-458.
[23] S. Minda, D. Brundage, Time differences in handwritten and computer documentation of nursing assessment. *Computers in Nursing* **12** (1993), 277-279.
[24] A.M. Hostgaard, C. Nohr, Dealing with organizational change when implementing EHR systems. *ud Health Technol Inform* **107** *(*2004), 631-634.
[25] M.V. Wideman, M.E. Whittler, M. Timothy. Barcode medication administration: Lessons learned from an intensive care unit implementation. *Advances in Patient Safety: From Research to Implementation,* 2005.
[26] M. Jebraeily, M. Ahmadi, A. Hajavi, M. Gohari, M. Sedghi, Z. Zareh, Electronic health records: personnel readiness assessment. *Journal of Health Administration* **13** (2010):17-24.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-226

How to Measure Physical Motion and the Impact of Individualized Feedback in the Field of Rehabilitation of Geriatric Trauma Patients

Amelie ALTENBUCHNER[a,1], Sonja HAUG[a], Rainer KRETSCHMER[b] and Karsten WEBER[a]
[a]*Institut für Sozialforschung und Technikfolgenabschätzung (IST), Ostbayerische Technische Hochschule Regensburg (OTH), Germany*
[b]*Klinik für Unfallmedizin, Caritas Krankenhaus St. Josef Regensburg, Germany*

Abstract. Background: This preparatory study accelerates an implementation of individualized monitoring and feedback of physical motion using conventional motion trackers in the rehabilitation process of geriatric trauma patients. Regaining mobility is accompanied with improved quality of life in persons of very advanced age recovering from fragility fractures. Objectives: Quantitative survey of regaining physical mobility provides recommendations for action on how to use motion trackers effectively in a clinical geriatric setting. Methods: Method mix of quantitative and qualitative interdisciplinary and mutual complementary research approaches (sociology, health research, philosophy/ethics, medical informatics, nursing science, gerontology and physical therapy). While validating motion tracker use in geriatric traumatology preliminary data are used to develop a target group oriented motion feedback. In addition measurement accuracy of a questionnaire about quality of life of multimorbid geriatric patients (FLQM) is tested. Conclusion: Implementing a new technology in a complex clinical setting needs to be based on a strong theoretical background but will not succeed without careful field testing.

Keywords. Health monitoring, motion tracker, quality of life, geriatric assessment

1. Introduction

Due to demographic changes in Germany, the amount of those in need of care grows [1], which entails a known phenomenon: the management of falls and injuries in the elderly is complicated by side conditions like dementia and care needs. The anticipated increase of patients with hip fractures is 70%, for people over 80 years even 150% in the next three decades [2]. Those who already suffered an injury have an increased risk of a following fracture [3]. Patients will need in-patient and surgical therapy [4]. Lack of specialists in Germany and growing costs in healthcare and nursing care demand efforts to implement innovative technology in these sectors [5]. Early postoperative mobilization is essential to prevent muscular atrophy and contractures [6]. A functional musculoskeletal system is needed to maintain autonomy. Early mobilization leads to a statistically significant earlier hospital discharge [7]. Also dementia patients benefit from

[1] Corresponding Author: Amelie Altenbuchner, Institut für Sozialforschung und Technikfolgenabschätzung (IST), Ostbayerische Technische Hochschule Regensburg (OTH), Seybothstraße 2, D-93053 Regensburg E-Mail: Amelie.Altenbuchner@oth-regensburg.de

exercise therapy [8, 9, 10]. Future research should focus on strategies how to continuously measure patient's motion, developing complex personalized interventions [10]. This claim accompanies a current desideratum of gerontology: Exploring practical devices for continuous vitality measurement in everyday conditions [1], as low costs and usability plead for clinical practice [11]. Only by addressing age-related challenges in the transfer from the laboratory into the field setting, acceptable usability of the device in the target population can be guaranteed. Evidence based research faces special methodological challenges [12, 13]: General findings about treatment response and rehabilitee success do not exactly fit individual needs. There is an issue about scientific objectivity in efficacy studies on industrial devices. In the geriatric setting (relatives, care levels, dementia) decision-making and communication of information is complicated.

Pending issues [11] on how to survey physical motion, provide feedback and measure quality of life in the rehabilitation process of geriatric trauma patients have to consider multiple perspectives: medicine, healthcare, patients, technical and social innovations, ethics and law. A preparatory study is conducted to answer a bundle of research questions and test methodological preliminary considerations concerning operationalization, data collection and analysis. This paper concentrates on the following research issues: How can customary motion trackers be used in hospitals? Are motion trackers accepted in the target group and why (not)?

2. Methods

Especially in the multidisciplinary and interdisciplinary field of geriatric medicine [14, 15], an effective development and implementation of technical and social innovations requires interdisciplinary research. Researcher and clinicians are have to develop a research framework using a method mix of quantitative and qualitative methods to increase validity of measured values [16]. Method triangulation reinforces the credibility of the qualitative findings in addition [17].

This preparatory study explores the research setting of a geriatric ward. Validity (content validity, construct validity, criterion validity) and reliability of questionnaires and measurement devices in this setting are tested. Based on the findings a motion feedback meeting the needs of the subjects is developed. The development of the instruments is based on a systematic literature review about quality of life in old age, the acceptance of technology and innovative assistant systems. Examining application of motion tracker and other instruments by statistical reliability and plausibility analysis and by qualitative techniques of fieldwork, to uncover possible error sources and to reinvent research goal based solutions. Therefore, writing a research record about all events after every research date interpreting based on inductive techniques as casuistry or as classifications, comparisons and contrasts, is necessary. Overall, particularly values and principles of research ethics are paid attention [18].

2.1. Subjects

The research design of the study (pretest-posttest-design) is based on a longitudinal data collection. Subjects are patients on the geriatric trauma ward CURA at the Caritas-Krankenhaus St. Josef, Regensburg, Germany after surgery for hip fracture. Recruitment will start postoperatively by obtaining informed consent. Data collection begins post-surgery until discharge (about 14 days per patient) and in the domestic situation after

rehabilitation center. Patients who suffer from severe cognitive impairment (e.g. dementia, delirium) are also included if informed consent is given by their custodians.

No medical intervention is part of the study. The device complements hitherto applied vitality and motion measurement in daily clinical routine. Additionally, the main study will have three four-month research periods to investigate how feedback affects motion and how wearing a device alone affects motion itself (nonequivalent-control-group-design). For each period, a number of n=48 is expected. Data collection includes relevant data from patient records like geriatric assessment scales of cognition, emotional status, mobility, status of sensory organs, care level, independence in everyday life, psychosocial aspects [19, 20].

The prospective study includes all hip fracture patients who agree to participate from 17/11/06 to 18/02/28 (n =10). These patients are provided with a motion tracker shortly after surgery. Data collection will cover the remaining inpatient hospital stay, usually 10 to 14 days. FLQM questionnaire [21] is also tested on other geriatric patients who are not included in motion tracking ($n \geq 30$) to run an analysis of reliability of the scales. First when starting and secondly after finishing the motion-tracking subjects are surveyed on their subjective quality of life and the acceptance of the applied technology. Next to testing the reliability of scales and their level of representing the research questions, the usability for the target group is reviewed. Based on the findings and on the literature review motion feedback will be developed for the group of old patients.

2.2. Questionnaires and interview questions

The prospective interest & methodological approach is how to measure subjective quality of life in very old aged patients who recently had hip fracture surgery as well as how to measure acceptance of innovative technology and wearing a tracking device for the first time. Measuring historically, culturally and societally determined subjective quality of life means that questioning should be held clear and short, especially in the case of old and very old people [22]. Unpleasant feelings should not be addressed and an actual change in the subjective quality of life should be possible to be determined [23].
FLQM [21] is an individualized instrument for the quality of life of multimorbid older people, comprising qualitative and standardized quantitative parts. FLQM was chosen based on a systematic literature review. Subjects get a printed table containing a scale (1-6, respectively 1-5) with written explanations in large font. The interviewer supplies data entry on an extra scale shed. Subjects themselves indicate four to six areas of life important to their individual quality of life. Subjects are rating each area according to their individual degrees of “contentment” and “importance”. In addition, two general statements on quality of life and the current emotional status are recorded. Furthermore, the FLQM includes statements of the Philadelphia Geriatric Centre Morale Scale (PGCMS). Reliability of scales will be analyzed in 18/02.

In the project BETAGT [24] a multimethod instrument for measuring acceptance of new technology in institutional contexts was developed that suggests a three-scale questionnaire for old people. For the prospective study, the measurement of acceptance is included in the measurement of quality of life that takes a relatively long amount of time. For this reason, the items of BETAGT have to be open interview questions, about subject’s previous experience, their attitude and knowledge about technology and new technology. Data analysis by thematic encoding is scheduled in 18/02.

2.3. Monitoring physical motion and patient records

The choice of the motion tracker (FitBit Alta HR) is based on the quantitative criteria "validity", "battery boost and dwell" as well as on the qualitative criteria "chafing prevention" and "water density" due to hygiene requirements. The prospective study measures physical motion and evaluates the usage of the device FitBit Alta HR. Battery life covers five days. Data downloading is possible while charging the device by connecting to a personal computer, laptop or a tablet or via Bluetooth using a smartphone application without recharging. Therefore, study personal has to see the patient on a regular basis which is an appropriate time to check the technical function, patient comfort and continuity of measurement. The field experiment also aims at finding out how charging and managing data retrieval. Visiting hours are from 8 am to 8 pm. An online tool of FitBit provides a graphical data review, depending on the individually set goal (preset in every FitBit: 10.000 steps per day, which is not usable for patients). The smallest time interval consists of 15 minutes. Heartrate is displayed in a three-colored systematic line chart. Extracting the data via .csv includes motion (steps on a daily basis and within a 15-minutes period) and sleeping activity.

To find out patient record data relevant for further analysis and the appropriate data collection process respecting data privacy issues, a conception of patient education including ethical and legal aspects was established. A manual data collection and a literature review on geriatric assessment scales is going on, which will lead into explorative data analysis with more extensive analysis.

2.4. Motion feedback

Development of motion feedback follows certain study findings [25, 26, 27, 28, 29] reviewed systematically by means of searching PUBMED on "exercise" AND "gerontology" AND "feedback" (filter: humans, last 10 years). For promotion of motion in geriatric rehabilitation a weekly feedback was given in numerically and graphically style. The displayed data based on the motion time in comparison to the goal time, whereas goals were adapted weekly [30].
Considering the fact that patients in different stages of cognitive impairment (e.g. dementia, delirium) are included in the study, findings and successful methods from practice projects using innovative technology have been considered [31, 32, 23, 34]: Using tablets was found to be a practicable medium to communicate with the patients and to activate them. However, light and visibility conditions as well as standardized, ritualized procedures with defined beginnings and ends have to be carefully defined. The motion feedback is based on these findings and will be put into test in January 2018.

3. Results

The prospective study detects errors and constructs solutions outside of the laboratory in the research field, so that reliable and valid measurement instruments are available for subsequent studies.

It turned out, that patients with visual disability cannot read their scale sheds of FLQM. So, the interviewer had to read the questions and scales out loud, sometimes two or three times, which led to a lack of clearness on grades of scales. This caused subjective requests of the interviewer. PGCMS was too tedious for subjects and addressed very

private emotions. Similar diagnostic data can be found in the scores of geriatric assessment, which is part of the medical treatment process anyway. Having answered the FLQM for about 30-45 minutes, only a few patients are able to answer additional questions concerning the acceptance of technology. They seem to remember using innovative technologies before on following days, when they feel regenerated.

Some problems and challenges summed up: Organizational reasons in hospitals do not allow a concurrent data collection from patient records and tracked data. A new subject can only be added every month in the FitBit system (age, weight, gender, height, step length, arm side can be changed regularly), so that a large number of devices is necessary. Number of steps are displayed only for divisions of 15-minute periods per hour. The device cannot always distinguish between steps and motion on ergo-bike or slowly driving a wheelchair. It would be important to differentiate between steps or motion on ergo-bike or driving a wheelchair. Quantitative analyzes are currently running, and will be published soon.

4. Conclusion

Using customary motion trackers as a clinical assessment equipment for motion, heartrate and sleeping/night activity offers an individualized, objective, efficient, continuous measurement on the actual progress of rehabilitants recovering from a hip fracture. Therefore, they help to assess the rehabilitation process, also after the hospital stay back in the home environment of the patients. If treatment advances are low, it is possible to intervene early (through feedback).

The great challenge is transferring study findings from the laboratory to the field and into everyday practice. Customary devices are not produced for research interest, but for the general user market. The elderly do accept the device as it wears like a wristwatch, however only a small minority will be able to operate the motion tracker without help.

A disadvantage for the quantitative researcher is the restricted access to the collected user data. For the implementation in the clinical and home rehabilitation process, relatives or caregivers have to be provided with a practicable solution how to interpret the resulting numbers and graphics. In this respect aim of the study is the development of a visual and audible output of motion (motion feedback) which can easily be used by patients, caregivers or relatives. Data downloading is possible while charging the device by connecting to a personal computer, laptop or a tablet or via Bluetooth using a smartphone application without recharging. Therefore, study personal has to see the patient on a regular basis which is an appropriate time to check the technical function, patient comfort and continuity of measurement.

The fact that battery life covers only about five days is a challenge itself. This includes a lot of charging during the research process, which is considered nearly impossible during a longitudinal study with a large number of subjects. So, we consider to use a similar comfortable device that lasts up to one year for the next study. Subjects can keep their device on their wrists during data downloading, then is necessary at least once a month, and during feedback.

To measure the social-science aspects the interviewer should be alone in the room with patient and always write a research record. Questions about psychosocial aspects should be fully taken from already existing geriatric assessment scores in the hospital, so geriatric patients don’t get too tired. As a result, it should be possible to conduct interviews on following days. This could also reduce item nonresponse. Interruptions

have to be noted in the research record. Interruptions can be added as a variable to test the statistically on changes in response behavior.

It is necessary to collect scores from medical records immediately after including a new subject.

All in all, a field study researching the subjects is possible by accurate preparation. The ongoing preparatory study reveals some parts where measurement errors can be prevented in further research. In following studies, it is required to use different devices.

5. References

[1] A. Barth, G. Doblhammer, Physische Mobilität und Gesundheit im Alter. Ansätze zur Reduktion von Pflegebedürftigkeit und Demenz in einer alternden Gesellschaft, in: Die transformative Macht der Demografie, T. Mayer, ed., Wiesbaden, 2017, pp. 207–244.

[2] R. Lohmann, K. Haid, U. Stöckle, M. Raschke, Epidemiologie und Perspektiven der Alterstraumatologie, Der Unfallchirurg 110(6) (2007), 553-60.

[3] C.M. Klotzbuecher, P.D. Ross, P.B. Landsman, T.A. Abbott, M. Berger, Patients with Prior Fractures Have an Increased Risk of Future Fractures: A Summary of the Literature and Statistical Synthesis, Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 15(4) (2000), 721–739.

[4] R. Kretschmer, M. Loher, C. Apfelbacher, M. Nerlich, Herausforderung Fragilitätsfrakturen – aus Sicht eines Alterstraumazentrums, Public Health Forum 25(2) (2017), 151–154.

[5] K. Weber, N. Kleine, F. Pallas, M.-R. Ulbricht, Internet of Things, Participatory Sensing, Wearables: Neue Formen der Datengewinnung, neue Anwendungen, neue normative Überlegungen, neuer Regulierungsbedarf?, Bonn, 2016.

[6] K. D. Hahn, K.H. Winker, Bedeutung und Möglichkeiten der postoperativen Behandlung nach Verletzungen., OP Journal 16 (2000), 160–163.

[7] S.J. Schaller, M. Anstey, M. Blobner, T. Edrich, S.D. Grabitz, I. Gradwohl-Matis, M. Heim, T. Houle, T. Kurth, N. Latronico, J. Lee, M.J. Meyer, T. Peponis, D. Talmor, G.C. Velmahos, K. Waak, J.M. Walz, R. Zafonte, M. Eikermann, Early, goal-directed mobilisation in the surgical intensive care unit: A randomised controlled trial, The Lancet 388(10052) (2016), 1377–1388.

[8] L. Clare, Rehabilitation for People Living with Dementia: a Practical Framework of Positive Support, PLoS medicine 14 (2017), e1002245.

[9] H. Bork, Rehabilitation nach hüft- und knieendoprothetischer Versorgung älterer Menschen, Orthopäde, 46(1) (2017), 69–77.

[10] K.S. McGilton, A.M. Davis, G. Naglie, N. Mahomed, J. Flannery, S. Jaglal, C. Cott & S. Stewart, Evaluation of Patient-Centered Rehabilitation Model Targeting Older Persons with a Hip Fracture, Including Those with Cognitive Impairment, BMC geriatrics 13 (2013), 136.

[11] B. Grimm, S. Bolink, Evaluating Physical Function and Activity in the Elderly Patient Using Wearable Motion Sensors, EFORT open reviews 1(5) (2016), 112–120.

[12] Nationale Akademie der Wissenschaften Leopoldina, acatech – Deutsche Akademie der Technikwissenschaften, Union der deutschen Akademien der Wissenschaften, Medizinische Versorgung im Alter - Welche Evidenz brauchen wir?, Halle, 2015.

[13] Fava G. A., Guidi J., Information Overload, the Patient and the Clinician, Psychotherapy and Psychosomatics 76(1) (2007), 1–3.

[14] R. Schulz, H.-W. Wahl, J.T. Matthews, A. de Vito Dabbs, S.R. Beach, S.J. Czaja, Advancing the Aging and Technology Agenda in Gerontology, The Gerontologist 55(5) (2015), 724–734.

[15] H. Brandenburg, Multi- und interdisziplinäre Perspektiven, in: Angewandte Gerontologie: Interventionen für ein gutes Altern in 100 Schlüsselbegriffen, H.-W. Wahl, ed., 2nd ed., Stuttgart, 2012, pp. 28–33.

[16] H. Mayer, Pflegeforschung anwenden. Elemente und Basiswissen für das Studium, Wien, 2015

[17] D.F. Polit, C.T. Beck, B.P. Hungler, Lehrbuch Pflegeforschung: Methodik Beurteilung und Anwendung, Bern, 2004.

[18] Rat für Sozial- und Wirtschaftsdaten (RatSWD), Forschungsethische Grundsätze und Prüfverfahren in den Sozial- und Wirtschaftswissenschaften, Berlin, 2017.

[19] Deutsche Gesellschaft für Unfallchirurgie, AltersTraumaRegister DGU. Datenerhebungsbogen 2017, Berlin, 2017.

[20] A.E. Stuck, Geriatrisches Assessment, in: Angewandte Gerontologie: Interventionen für ein gutes Altern in 100 Schlüsselbegriffen, H.-W. Wahl, ed., 2nd ed., Stuttgart, 2012, pp. 116–121.
[21] M. Holzhausen, Lebensqualität multimorbider älterer Menschen: Konstruktion eines neuen individualisierten Messverfahrens, Bern, 2009.
[22] G. Netuveli, D. Blane, Quality of Life in Older Ages, British medical bulletin 85(1) (2008), 113–126.
[23] A. Gupta, Assessmentinstrumente für alte Menschen: Pflege- und Versorgungsbedarf systematisch einschätzen, Bern, 2012.
[24] K. Claßen, Technik im Alltag, in: Angewandte Gerontologie: Interventionen für ein gutes Altern in 100 Schlüsselbegriffen, H.-W. Wahl, ed., 2nd ed., Stuttgart, 2012, pp. 499–506.
[25] K. Broekhuizen, J. de Gelder, C.A. Wijsman, L.W. Wijsman, R.G.J. Westendorp, E. Verhagen, P.E. Slagboom, A.J. de Craen, W. van Mechelen, D. van Heemst, F. van der Ouderaa, S.P. Mooijaart, An Internet-Based Physical Activity Intervention to Improve Quality of Life of Inactive Older Adults: a Randomized Controlled Trial, Journal of medical Internet research 18(4) (2016), e74.
[26] D.E. Forman, K. LaFond, T. Panch, K. Allsup, K. Manning, J. Sattelmair, Utility and Efficacy of a Smartphone Application to Enhance the Learning and Behavior Goals of Traditional Cardiac Rehabilitation: a Feasibility Study, Journal of cardiopulmonary rehabilitation and prevention 34(5) (2014), 327–334.
[27] N. Lauzon, C. B. Chan, A. M. Myers, C. Tudor-Locke, Participant Experiences in a Workplace Pedometer-Based Physical Activity Program. Journal of Physical Activity and Health, 5(5), 2008, 675–687.
[28] B.J. Nicklas, J.E. Gaukstern, K.M. Beavers, J.C. Newman, X. Leng, W.J. Rejeski, Self-Monitoring of Spontaneous Physical Activity and Sedentary Behavior to Prevent Weight Regain in Older Adults, Obesity 22(6) (2014), 1406–1412.
[29] F. Ofli, G. Kurillo, Š. Obdržálek, R. Bajcsy, H.B. Jimison, M. Pavel, Design and Evaluation of an Interactive Exercise Coaching System for Older Adults: Lessons Learned, IEEE journal of biomedical and health informatics 20(1) (2016), 201–212.
[30] N.M. Peel, S.K. Paul, I.D. Cameron, M. Crotty, S.E. Kurrle, L.C. Gray, Promoting Activity in Geriatric Rehabilitation: a Randomized Controlled Trial of Accelerometry, PloS one 11(8) (2016), e0160906.
[31] A.J. Astell, M.P. Ellis, L. Bernardi, N. Alm, R. Dye, G. Gowans, J. Campbell, Using a touch screen computer to support relationships between people with dementia and caregivers, Interacting with Computers 22(4) (2010), 267–275.
[32] F.S. Lim, T. Wallace, M.A. Luszcz, K.J. Reynolds, Usability of Tablet Computers by People with Early-Stage Dementia, Gerontology 59(2) (2013), 174–182.
[33] J. Nordheim, S. Hamm, A. Kuhlmey, R. Suhr, Tablet-PC und ihr Nutzen für demenzerkrankte Heimbewohner: Ergebnisse einer qualitativen Pilotstudie, Zeitschrift für Gerontologie und Geriatrie 48(6) (2015), 543–549.
[34] D. Upton, P. Upton, T. Jones, K. Jutlla, D. Brooker, Evaluation of the Impact of Touch Screen Technology on People with Dementia and their Carers Within Care Home Settings, Project Report, Worcester, 2011.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-233

A Survey of Managers' Access to Key Performance Indicators via HIS: The Case of Iranian Teaching Hospitals

Maryam AHMADI[a], Farid KHORRAMI [a,1,] Afsaneh DEHNAD[b], Mohammad H GOLCHIN[c], Mohsen AZAD[d], Shafei RAHIMI[d]
[a] *Dept. of Health Information Management, School of health management and information sciences, Iran University of Medical Sciences, Tehran, Iran*
[b] *Dept. of English Education, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran*
[c] *Dept. Of Nursing Office, Shahid Mohammadi Hospital, Hormozgan University of Medical Science, Bandar Abbas, Iran*
[d] *Cardiovascular Research Center, Hormozgan University of Medical Sciences, Bandar Abbas, Iran.*

Abstract. Background: The challenges of using health information systems in developing countries are different from developed countries for various reasons such as infrastructure and data culture of organizations. Objective: The aim of this study is to assess managers' access to key performance indicators (KPI) via Hospital Information System (HIS) in teaching hospitals of Iran. Methods: All managers (Census method) of the four teaching hospitals affiliated to Hormozgan University of Medical Sciences (HUMS) were included in this study. KPIs which are linked to the strategic objectives of organizations were adopted from the strategic plan of HUMS. The questionnaire used in this study included three categories: Financial, Human Resources and clinical. One-sample t-test was used and the significant difference score was calculated for the acceptable level. Results: We found that HIS cannot facilitate access to KPIs for managers in the main categories, but it was effective in two subcategories of income (p=0.314) and salary (P=0.289)). Conclusion: A study of barriers to the use of managers of HIS in hospitals is suggested.

Keywords. Hospital Information Systems, Health Information Management, Hospital Administration, Iran.

1. Introduction

The benefits of the adopting HIS in health care organizations, have been confirmed in several studies [1-6]. The data which was generated and transformed to information and knowledge by these systems highlight the vital role of information systems in an organization, i.e. information alone is not “power”, but using it at the right time and right place shows “power” [7]. Facilitation and acceleration of access to Key Performance Indicators (KPIs) are the main requirements of any organization after the implementation of the systems in hospitals [8]. Meeting its strategic goals, an organization needs to define

[1] Corresponding Author: Farid Khorrami, School of Health Management and Information Sciences, Iran University of Medical Sciences, Address: No. 6, Rashid Yasemi St., Vali-e-asr Ave., 1995614111 Tehran, Iran. Tel.: +98 9171982334; fax: +98 21 88883334 , E-mail: khorrami.farid@gmail.com

success and track its progress. KPI has an important role in determining the appropriateness of the main tools used to systematically monitor, evaluate, and continuously improve service performance in most advanced economies, in middle- and low-income countries [9].

Moreover, only information systems which potentially have the characteristics of accessibility, usability, reliability, adaptability and response time are capable of producing high quality data and information for a knowledge-based management in health care organizations [10]. Lammintakanen et al have categorized the barriers of information systems use into two groups:

- Barriers to use related to information content
- Barriers to use related to information systems [11]

However, the use of the information by managers has been reported to be the most important issue in organizations with a poor data culture. Focusing on the relationship between culture and use of information by the managers, Kevinen et al. recommended the application of different strategies to strengthen the use of information in organizations [12]. As a result, when evaluating and using the system, managers should take into consideration the usability of the information [10,13]. The purpose of usability evaluation is to determine the strengths and weaknesses of information systems and to provide guidelines for the improvement of their applications [10,14,15].

The past decade has witnessed the rapid adoption of HISs in many hospitals in Iran. Although extensive research has been carried out on the evaluation of HISs, no single study exists which adequately covers the use of the information generated by these information systems in different areas of hospital management. The aim of this study is to assess of managers' access to KPI via HIS in teaching hospitals.

2. Methods

This descriptive and analytical cross-sectional study was conducted in 2016. The study population included all strategic, tactical and operational managers in four teaching hospitals of Hormozgan University of Medical Sciences (HUMS). Due to limited research population, census method instead of sampling was used in this study.

Since KPIs are linked to the strategic objectives of organizations [9,16], KPIs of this study were extracted from the strategic plan of HUMS and the national health indicators published by Iran's Ministry of Health and Medical Education [17]. Considering the HIS adoption and implementation, the studied hospitals were at different stages. Hence, hospital managers answered questions regarding accessibility to KPIs by HIS according to the stage of HIS adoption. The questionnaire included three categories: Financial (27 items), Human Resources (19 items) and clinical (22 items).Questionnaires were distributed on the basis of organizational positions. They were asked to identify access to the information with three-choices of answers "yes" (one point) and "no" and "I have no idea" (zero points). To calculate the score of each item, the scores were added together and divided by the number of questions. Then, the score for each section was assigned between zero and one, and scores above 0.7 were considered to be acceptable. The significant difference score was calculated for the acceptable level and one-sample t-test was used. ($\mu \neq 0.7$vs.$\mu = 0.7$)

3. Results

A total of 126 questionnaires, completed by 106 managers, were collected. Some of the questionnaires were completed by the same managers in different areas for example the manager of financial office completed two more questionnaires in addition to the questionnaire directly related to his own area. The participants of the study consisted of 81.9% female and 18.1% male managers. In addition, 79% of the managers under the study had bachelor's degree with positions mostly at tactical and operational management level while strategic managers formed 4.8% of the total research population were general practitioners and specialists.

To assess of managers' access to KPIs via HIS, the areas under the study were divided into 6 subcategories including 37 indicators. According to the results (Table 1), managers' access to financial information via the HIS was significantly lower than the appropriate level (P=0.004). In this category, access to information related to income, was at an appropriate level (p=0.314) whereas the subcategory of access to the information related to costs was significantly lower than the appropriate level (p<0.001).

Managers' access to human resources management indicators via the HIS was significantly lower than the appropriate level (P=0.003). Moreover, in the subcategories, access to information related to productivity was significantly lower than the appropriate level (P=0.001). However, the information related to salary was in the appropriate level (P=0.289). Managers' access to clinical information via HIS was significantly lower than the appropriate level (p<0.001) and in the subcategories, the average scores were significantly below the appropriate level (P <0.001). On the other hand, the rate of managers' access to KPIs through HIS in financial, human resources and clinical management were (56%), (48%) and (47%), respectively. These findings appeared to be lower than the appropriate level (Table 1).

4. Discussion

The main purpose of this study was to assess of managers' access to KPIs via HIS in HUMS hospitals. Although the findings of this study suggest that HIS in hospitals under the study did not facilitate managers' access to information in any of the main categories while in some subcategories (income and salary) they were favorable. Similarly, other studies in Iran, reported of inefficiency of HISs in meeting the information needs of managers too [14,18-21].

Table 1. Results of assess of managers' access to KPIs via HIS in HUMS

Categories	Subcategories	Mean scores	Standard deviation	t_0	95% Confidence Interval	P-value
Financial	Cost	0.37	0.23	-6.02	(0.26-0.49)	<0.001
	Income	0.75	0.18	1.06	(0.65-0.84)	0.314
	Total	0.56	0.28	-3.05	(0.46-0.65)	0.004
Human Resources Management	Productivity	0.36	0.24	-4.86	(0.21-0.51)	0.001
	Salary	0.59	0.36	-1.09	(0.37-0.81)	0.298
	Total	0.48	0.33	-3.36	(0.34-0.61)	0.003
Clinical	Structure/Process	0.54	0.26	-5.63	(0.48-0.59)	<0.001
	Outcome	0.41	0.24	-10.59	(0.35-0.46)	<0.001
	Total	0.47	0.26	-10.94	(0.43-0.51)	<0.001

According to research findings, managers' access to KPIs showed that the financial category appeared to be relatively more effective (56%) than other categories. These finding supports the findings of other Iranian researchers in universities of medical sciences in recent decades [14,18-21]. This finding places Iran in the early level of computerization on the basis of the categories of Medical Records Institute [8]. Moreover, most HIS projects in Iran are planned for automating financial processes in hospitals [18,21,22] therefore, it goes without saying that managers' access to financial information is more than other categories especially more than the main category of clinical information. Low level of access to subcategory of the information related to cost was an indirect issue that in which most of the information related to cost of the hospital were not recorded by the system. Often, other information systems which are not integrated with the HIS and are independent of the main system of the hospital provide the access to the information related to the cost for the managers. While an integrated HIS including human resources systems, data management, and financial systems and health care processes are more effective in increasing the use of information systems [12,23,24].

Research findings in the category of human resources show that HIS was not favorable in this category however the effect of HIS was desirable in the subcategory of information related to employees' salary. This subcategory is one of the administrative uses of HIS which has been stressed on its adoption in Iran [18,21,22,25]. It is worth mentioning that systems similar to this subcategory which are independent of HIS are extensively employed in hospitals but in most cases, they are not integrated with HIS.

Regarding managers' access to information in the clinical category, the finding of this study was not favorable and this is in line with the findings of other studies performed in Iran [14,18,19,21,22,25] and other developing countries [4,24,26,27]. This finding suggests that the effect of HIS on managers' access to information was reported as "unsatisfactory". There are two reasons which seem to be responsible for this. First, HIS is not function as a tool to produce the information required in the hospital [11]. In the development phase, weaknesses in various stages of development of information systems such as the analysis and design of the software caused by lack of needs assessment, engagement of mangers in adopting HIS due to problems in user-friendliness of systems, and imprecise and impractical reports on the functionality of systems [10,11,21,28] are some of the problems. In the startup phase, inadequate training of staff and users, the complexity of the system, lack of integration of information systems, lack of resources, inadequate understanding of information technology, poor support of the vendors to build a complete report development, absence of a long-term planning system, resistance to change are other reasons responsible for the weaknesses [10-12,18,21]. Second, information may be available in HIS but is not considered to support management for reasons such as lack of updated information in the system, non-valid information and information segregation in HIS, all of which may make the poor usability of information systems [11].

5. Conclusions

Due to the flaws of the system to meet the information needs of executives in different categories, a study of barriers to the use of information system managers in any organization is suggested. Moreover, at the time of purchase and procurement of information systems in the request for proposal, the information needs of clinical and

human resources categories should be identified, included and data integration policy should be followed up in HIS.

6. References

[1] N. Zakaria and S.A. Mohd Yusof, Understanding Technology and People Issues in Hospital Information System (HIS) Adoption: Case study of a tertiary hospital in Malaysia, J Infect Public Health 9 (2016), 774-780.

[2] S. Yoo, H. Hwang, and S. Jheon, Hospital information systems: experience at the fully digitized Seoul National University Bundang Hospital, J Thorac Dis 8 (2016), S637-641.

[3] R. Khajouei, N. Peek, P.C. Wierenga, M.J. Kersten, and M.W. Jaspers, Effect of predefined order sets and usability problems on efficiency of computerized medication ordering, Int J Med Inform 79 (2010), 690-698.

[4] J. Lee and B. Dowd, Effect of health information technology expenditure on patient level cost, Healthc Inform Res 19 (2013), 215-221.

[5] J.M. Han, Y.M. Chae, E.H. Boo, J.A. Kim, S.J. Yoon, and S.W. Kim, Performance analysis of hospital information system of the national health insurance corporation ilsan hospital, Healthc Inform Res 18 (2012), 208-214.

[6] C.J. Black AD, Pagliari C, Anandan C, Cresswell K, Bokun T, et al, The Impact of eHealth on the Quality and Safety of Health Care: A Systematic Overview, PLoS Med 8 (2011), e1000387.

[7] L. Farmer, Information competency and e-learning, IGI Global, United States of America, 2012.

[8] K.A. Wager, F.W. Lee, and J.P. Glaser, Health Care Information Systems A Practical Approach for Health Care Management, John Wiley & Sons, San Francisco, 2013.

[9] S. Rozner, Developing and Using Key Performance Indicators A Toolkit for Health Sector Managers, in: A.A.I. Health Finance & Governance Project, ed., United States Agency for International Development, Bethesda MD, 2013, p. 13.

[10] N. Urbach and B. Müller, The Updated DeLone and McLean Model of Information Systems Success, in: Information Systems Theory: Explaining and Predicting Our Digital Society, Vol. 1, Y.K. Dwivedi, M.R. Wade, and S.L. Schneberger, eds., Springer New York, New York, NY, 2012, pp. 1-18.

[11] J. Lammintakanen, K. Saranto, and T. Kivinen, Use of electronic information systems in nursing management, Int J Med Inform 79 (2010), 324-331.

[12] T. Kivinen and J. Lammintakanen, The success of a management information system in health care – A case study from Finland, Int J Med Inform 82 (2012), 90-97.

[13] M.M. Yusof, R.J. Paul, and L.K. Stergioulas, Towards a Framework for Health Information Systems Evaluation, in: Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), 2006, pp. 95a-95a.

[14] R. Safdari, H. Dargahi, L. Shahmoradi, and A. Farzaneh Nejad, Comparing four softwares based on ISO 9241 part 10, J Med Syst 36 (2012), 2787-2793.

[15] R. Ologeanu-Taddei, D. Morquin, and R. Bourret, Understanding the Perceived Usefulness and the Ease of Use of a Hospital Information System : the case of a French University Hospital, Stud Health Technol Inform 210 (2015), 531-535.

[16] K.W. Cho, S.-K. Bae, J.-H. Ryu, K.N. Kim, C.-H. An, and Y.M. Chae, Performance evaluation of public hospital information systems by the information system success model, Healthc Inform Res 21 (2015), 43-48.

[17] National Health indicators, in, Statistics and Information Technology Office, Iran, 2014, pp. 19-45.

[18] L. Ahmadian, S.S. Nejad, and R. Khajouei, Evaluation methods used on health information systems (HISs) in Iran and the effects of HISs on Iranian healthcare: a systematic review, Int J Med Inform 84 (2015), 444-453.

[19] M.A.A. Moghadam and A. Fayaz-Bakhsh, Hospital Information System Utilization in Iran: a Qualitative Study, Acta Med Iran 52 (2014), 855-859.

[20] F. Rangraz Jeddi, F. Abazari, A. Moravveji, and M. Nadjafi, Evaluating the ability of hospital information systems to establish evidence-based medicine in Iran, J Med Syst 37 (2013), 9904.

[21] F. Sadoughi, K. Kimiafar, M. Ahmadi, and M.T. Shakeri, Determining of Factors Influencing the Success and Failure of Hospital Information System and Their Evaluation Methods: A Systematic Review, Iran Red Crescent Med J 15 (2013), e11716.

[22] H. Ayatollahi, Z. Nazemi, and H. Haghani, Patient Accounting Systems: Are They Fit with the Users' Requirements?, Healthc Inform Res 22 (2016), 3-10.

[23] A. Ismail, A.T. Jamil, A.F.A. Rahman, J. Madihah, A. Bakar, and N.M. Saad, The Implementation of Hospital Information System (HIS) in Tertiary Hospitals in Malaysia : A Qualitative Study, Malaysian Journal of Public Health Medicine 10 (2010), 16-24.
[24] H.H. Kim, K.-W. Cho, H.S. Kim, J.-S. Kim, J.H. Kim, S.P. Han, C.B. Park, S. Kim, and Y.M. Chae, New integrated information system for pusan national university hospital, Healthc Inform Res 17 (2011), 67-75.
[25] M. Abdekhoda, M. Ahmadi, M. Gohari, and A. Noruzi, The effects of organizational contextual factors on physicians' attitude toward adoption of Electronic Medical Records, Journal of Biomedical Informatics 53 (2015), 174-179.
[26] B. Tilahun and F. Fritz, Comprehensive evaluation of electronic medical record system use and user satisfaction at five low-resource setting hospitals in ethiopia, JMIR Med Inform 3 (2015), e22.
[27] A. Boonstra and M. Broekhuis, Barriers to the acceptance of electronic medical records by physicians from systematic review to taxonomy and interventions, BMC Health Serv Res 10 (2010), 231.
[28] B. Rahimi, R. Safdari, and M. Jebraeily, Development of Hospital Information Systems: User Participation and Factors Affecting It, Acta Informatica Medica 22 (2014), 398-401.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-239

Towards Designing a Secure Exchange Platform for Diabetes Monitoring and Therapy

Jan-Patrick WEIß[a,1], Tobias WELZEL[b], Bernd J. Hartmann[b], Ursula HÜBNER[a] and Frank TEUTEBERG[c]
[a]*Health Informatics Research Group, Osnabrück University of AS, Germany*
[b]Institute of Local Government Law & Public Admin., University Osnabrück, Germany
[c]Research Group Accounting & Information Systems, University Osnabrück, Germany

Abstract. Background: Diabetes mellitus is one of the most prominent examples of chronic conditions that requires an active patient self-management and a network of specialists. Objectives: The aim of this study was to analyze the user and legal requirements and develop a rough technology concept for a secure and patient-centered exchange platform Methods: To this end, 14 experts representing different stakeholders were interviewed and took part in group discussions at three workshops, the pertinent literature and legal texts were analyzed. Results: The user requirements embraced a comprehensive set of use cases and the demand for "one platform for all" which is underlined by the right for data portability according to new regulations. In order to meet these requirements a distributed ledger technology was proposed. Conclusion: We will therefore focus on a patient-centered application that showcases self-management and exchange with health specialists.

Keywords. Diabetes mellitus, patient-centered care, distributed systems.

1. Introduction

Diabetes mellitus has become a serious public health problem in many countries around the globe that affects 1 in 11 adults, with increasing prevalence values worldwide [1]. 90% of all diabetes patients have type 2 (DM II) [1]. Often, it is associated with co-morbidities particularly the metabolic syndrome, which includes obesity, dyslipidemia and arterial high blood pressure [2]. Diabetic patients including those with DM II possess a high risk of developing late complications such as the diabetic foot syndrome and retinopathy [3]. The severity, chronicity and complexity of DM II leads to the question of how the condition can be improved and of how health IT can help to solve challenges specific to DM II. These are (1) the early diagnosis of the disease, (2) prevention of late consequences, (3) inter-professional cooperation (e.g. medical specialties, wound experts, physiotherapists), (4) monitoring of crucial blood values to adjust the therapeutic regimen, (5) decision support with respect to the appropriate therapy, (6) self-management of the patient (patient as expert of his/her body) and (7) detection of complications. The importance and consequences of the disease also entail the question

[1] Corresponding Author: Jan-Patrick Weiß, Osnabrück University of AS, Health Informatics Research Group, PO Box 1940, 49009 Osnabrück, Germany; E-Mail: j.p.weiss@hs-osnabrueck.de

of how routinely compiled electronic patient data can be used to advance the knowledge in health services and precision medicine.

Due to the high potential of health IT for managing patients with diabetes, a wealth of systems has been developed early on to help supporting patients and health professionals. It started with smart devices for measuring blood glucose levels in particular for continuous measurements and for delivering insulin [4,5]. These devices that were initially standalone are getting more and more integrated into electronic health record (EHR) systems [6] or are connected to apps on mobile phones (e.g. [7]).

EHR systems have become focal points for improving the screening of prediabetic persons or those at risk [8] and also have included decision support functions that integrate guidelines and patient data [9]. Improved decision making and lower longtime blood glucose levels in patients could be demonstrated due to EHR implementation and use by physicians [10]. At system level, the utilization of an EHR in an integrated practice network resulted in higher quality of care for diabetic patients. The EHR has been employed to manage the delivery of a multi-faceted bundle of best care interventions of multidisciplinary teams. The conditions of the patients improved significantly [11]. In addition, comparable EHRs in different health provider sites allowed performance benchmarks to be conducted as has been shown for the care of diabetic patients [12].

Shared EHR systems allow e-consultations between primary physicians and specialists as they are recommended for diabetic patients with high longtime blood glucose values or hypoglycemia [13]. EHRs are also suitable for dedicated analyses of large data sets, for example of longitudinal patient data to model disease progression also on a personalized level [14]. The ultimate vision is to make use of omics data integrated in an EHR with decision support elements to offer personalized treatments for DM patients [15].

Although there is an impressive track record of electronic systems for use in the diagnosis and therapy of diabetes, problems emerge the more types of different professions and institutions access the data. Practical implementations of shared EHR systems for inter-professional use at the interface between primary and secondary care demonstrated the feasibility of implementation but also highlighted difficulties, e.g. regarding IT governance [16], or legal and security challenges that are to be solved beforehand in particular with respect to allocation of responsibility, documentation routines and access control [17].

While blood glucose monitoring systems have become state of the art and are increasingly used, in the majority EHR projects are still lighthouse initiatives in many countries simply because EHRs are not ubiquitous available or because there is no infrastructure to exchange data securely. An Austrian study showed that diabetic care is highly fragmented with 90% of the patients receiving care by more than one provider. The authors concluded that a shared electronical record, like the Austrian ELGA, would be the most suitable instrument to build information continuity for diabetic patients [18].

This overview, which highlights the state of the art, the potentials and the challenges, indicates the need for further research on how to design, implement and deploy an electronic platform that could serve as a regional exchange and networking instrument dedicated to improving the care of diabetic patients. The aim of this paper therefore is to present the results of requirements engineering, analysis of legal requirements, and a rough technology concept for the implementation of such an exchange platform.

2. Methods

In order to integrate the scientific knowledge base as well as knowledge from the application domain, the development of the concept presented here follows the Design Science approach according to Hevner et al. [19]. As part of an iterative development and evaluation process, the steps requirements analysis, design, implementation and evaluation are to be carried out in several rounds. The project is currently in the requirements analysis and systems design phase. Up to now, a literature search [20] and a qualitative cross-sectional analysis of workshop results (unstructured expert interviews) were performed based on [21,22]. In the period from May 2017 to September 2017 a total of 3 workshops were held, each with a duration of 4 hours and consisting of 4 experts, the first among physician and laboratory experts, then among pharmacists and a lawyer, and finally among IT experts, nurses and patients. In addition, 14 individual experts (4 physicians, 2 laboratory experts, 4 pharmacists, 1 lawyer, 1 IT-expert, 1 nurse, 1 user) interviews were conducted. The interview duration was approximately 2 hours each. An interview guideline was used during the workshops and individual interviews. The interview guideline consisted of open questions about the therapy and care of diabetic patients and how these processes could be supported by health IT. Accordingly, the following central open questions were discussed: What are the current problem areas in diabetes therapy? Which systems, apps and devices are used? How is their acceptance? Who exchanges which data with whom? How could medication management be implemented? Where are legal obstacles (What must be considered legally? data protection, data security?)? Which stakeholders are involved in diabetes therapy and how? Where could IT help to simplify the processes of diabetes therapy? What functionalities should a platform have that connects the stakeholders involved in diabetes therapy? Which data can be stored in a digital health record?

3. Results

3.1. User Requirements

Table 1 summarizes the results of the requirements analysis according to the different targets user, process, application and utilization. It also shows the source of the individual requirements. One of the central requirements derived from these various sources was the need for “one platform for all” including for the patient. Therefore, the requirements were grouped and conceptualized in graphical form following a layer scheme that represents data management, the applications, the users, and presentation and interaction (Fig. 1). This diagram also comprises dynamic information about different workflows from admission/discharge (hospital view) to rehabilitation and consultation (hospital – general practitioner or specialist – general practitioner view). It also reveals the desirable flows of information, medication, money, and prescriptions. The digital patient record system in the data management layer symbolizes the need to integrate the data. However, it does not automatically imply a centralized way of storing the data but can also be regarded as a repository that keeps track of the documents and data stored for each patient at local sites. The requirements shown in Fig. 1 reflect the combined set of all requirements without prioritizing certain groups of users, applications, workflows, and data. The resulting picture, therefore, is complex and mirrors an ideal situation rather than the design of a specific dedicated application system.

Table 1. Preliminary requirements summary through an initial workshop and literature research

	Requirement	Description	R
User	Patient-centered care / patient participation	It must be ensured at all times that the patient retains control over his/her data, that data is only used after its release and that the patient can view his/her data at any time.	E
	One platform for all	Digital networking of all stakeholders involved in patient care in the use case of diabetes therapy and virtual diagnostics and creation of adequate organizational structures and low-threshold processes.	W
Process	Complement and not replace	Telemedicine as a supplement to personal doctor/pharmacist patient contact	E
	Legal certainty in the specification and use of telemedicine measures	Assessment of the legal framework conditions for the use of telemedicine measures, in particular in the area of data protection and medical device law, in order to provide actors with security of action.	L [23]
Application	Ease of use	Faster access to relevant information (e. g. medication, therapy courses, interactions, diseases, etc.) in emergency situations	E
	Information quality	Holistic clinical picture due to large database (if data sharing is approved by the patient)	W
Utilization	System for improvements	Telemonitoring to improve the course of therapy (monitoring/compliance); anonymous analyses of patient data for decision support (pattern recognition; creation of new forms of therapy)	L [24,25]
	Reminder functions	Reminder of taking medication - it is conceivable that this could be linked to wearables such as a smart watch, which the patient always carries with him/her. Reminder of regular measurements of certain bodily functions with subsequent automatic transmission of the data to the doctor and pharmacist.	L [26]
	Documentation function	Nurses document the patient's state of health by means of photographs and/or parameters to be determined. Patients document their state of health, for example by regularly answering various questions about their state of health.	E
	Consent form	Possibility of documenting consent to collecting, processing and using personal data	E
	Accounting	Support of documentation, billing and application to insurance providers	E

Annotation: R = Research method, E = Expert interview, L = Literature review, W = Workshop

3.2. Legal Requirements

Due to the comprehensiveness of the “one platform for all”, the legal requirements need to be focused in particular. In Germany, the development of eHealth software, applications and devices is governed by a wide range of legal requirements. For the EHR system described above the Medizinproduktegesetz (MPG, Medical Devices Act) [27] is among the relevant laws. Section 3 no. 1 MPG states that the MPG is applicable to software, if the computer program serves one of the purposes listed in Section 3 no. 1 lit. a-d MPG (examples are the recognition, prevention, monitoring, or treatment of diseases). Whether software fulfills one of these purposes depends at heart on the manufacturer’s instructions, as Section 3 no. 10, 15 MPG states. In order to avoid bypassing the strict regulations of the MPG, the law provides prohibition of arbitrariness: If the software’s objective purpose is for medical use only, it falls under the regulations of the MPG, divergent manufacturer’s instructions notwithstanding [28,29]. Therefore, software which serves as diagnostic support (e.g. as tool for blood glucose or medication

monitoring) or compares laboratory results with reference values has to fulfill the MPG's legal requirements [29-31]. On the other hand, software that only serves as a tool for storing, reproducing, or transferring data is exempt from the MPG [29,30]. The MPG requires a CE certification (cf. Section 6 MPG), and the manufacturer in charge according to Section 5 MPG generally has to provide a clinical study before the product is allowed to launch (Sections 19 et seq. MPG). A violation of these provisions can be fined or may lead to criminal charges according to Sections 40 et seq. MPG. Additionally, the Medizinprodukte-Betreiberverordnung (MPBetreibV; Order on Medical Products' Operators) and the Heilmittelwerbegesetz (HWG, Law on Medicine Advertisement) are stating further obligations for the manufacturer [30,32].

Using personal data falls under the regulation of data protection laws. Currently the area of data protection in Germany is highly fragmented legally, containing at least six different data protection laws which are applicable to eHealth devices and software. Examples are the federal Bundesdatenschutzgesetz (BDSG; Federal Data Protection Act), Sozialgesetzbuch (SGB; Social Security Code) I, V, X, Telemediengesetz (TMG; Telecommunication Media Act) and on state level data protection laws and hospital laws. This situation will change on May 25th 2018, when the European Union's General Data Protection Regulation (GDPR) will become applicable. The GDPR's aim is to unify the protection of personal data in the EU [33]. As a regulation, the GDPR does not need to be transferred into national law, but is directly applicable in all member states [34]. The GDPR will bring about remarkable changes to data protection in Germany. On the one hand, GDPR will improve the users' rights to control their personal data (compared to the currently applicable BDSG [35,36]. Existing rights will be expanded. A visible example is the right to erasure (Article 17 GDPR). Currently, Section 35 BDSG contains a right to erasure only for specific data, whereas Article 17 GDPR expands this right to all personal data in special cases and thus integrates "the right to be forgotten" developed by the European Court of Justice [37]. Unprecedented in some respects are the right of access by the data subject (Article 15 GDPR) and the right to data portability (Article 20 GDPR). Developers of data processing technology are also obligated to protect data by design and by default (Article 25 GDPR). Violations are fined pursuant to Article 83 GDPR. On the other hand, GDPR does not live up fully to its own aims as only a few companies are obliged to install an independently acting data protection officer (Articles 37-40 GDPR) [38] and many regulations are incomplete and need to be filled up through national law. This may lead to fragmentation of data protection in the EU.

Further legal issues that need to be considered when developing and using software for medical purposes are the regulations of malpractice. In addition, the digitalization of other branches in the health sector leads to various issues with pharmaceutical law, as the example of a so called "video-pharmacy" in the German town of Hüffenhardt demonstrates [39,40]. The further activities will consist of a thoroughly legal study on the issues mentioned above and thus sets out to provide guidelines for designing EHR systems in accordance with the legal requirements described.

3.3. General Technical Concept

These complex requirements for "one platform for all" including the new legal obligations call for a highly secure backbone system. We therefore decided to employ distributed ledger technology. Data protection and data security is provided by the inherent mechanisms of distributed ledgers [41]. The technical benefit of distributed ledgers is the grouping of data blocks by combining data records into blocks and

guaranteeing data integrity using cryptographic methods [41]. One example for a type of data structure of distributed ledgers are blockchains. In blockchains new transactions append a new data block to the chain and each time it contains a checksum, i.e. a cryptographic hash function ("encryption"), which contains both the checksum of the previous block and that of the entire chain. This means that manipulation of a dataset could be detected. With this structure, the current status of a care record depends on all past records. The patient record becomes a single series of historical care events, regardless where he/she was treated. Furthermore, the use of cryptographic hash functions help to ensure the regulation of access rights by patients, to track their own data within the health care system. In principle, HL7 FHIR and SNOMED CT would be ideal to ensure interoperability [42]. The concept therefore builds on these components. In practice, both are rarely used, and dedicated interfaces will have to be developed for exchanging patient data. A distributed ledger records these exchanges as transactions with references to the respective patient data and therefore serves as an audit trail.

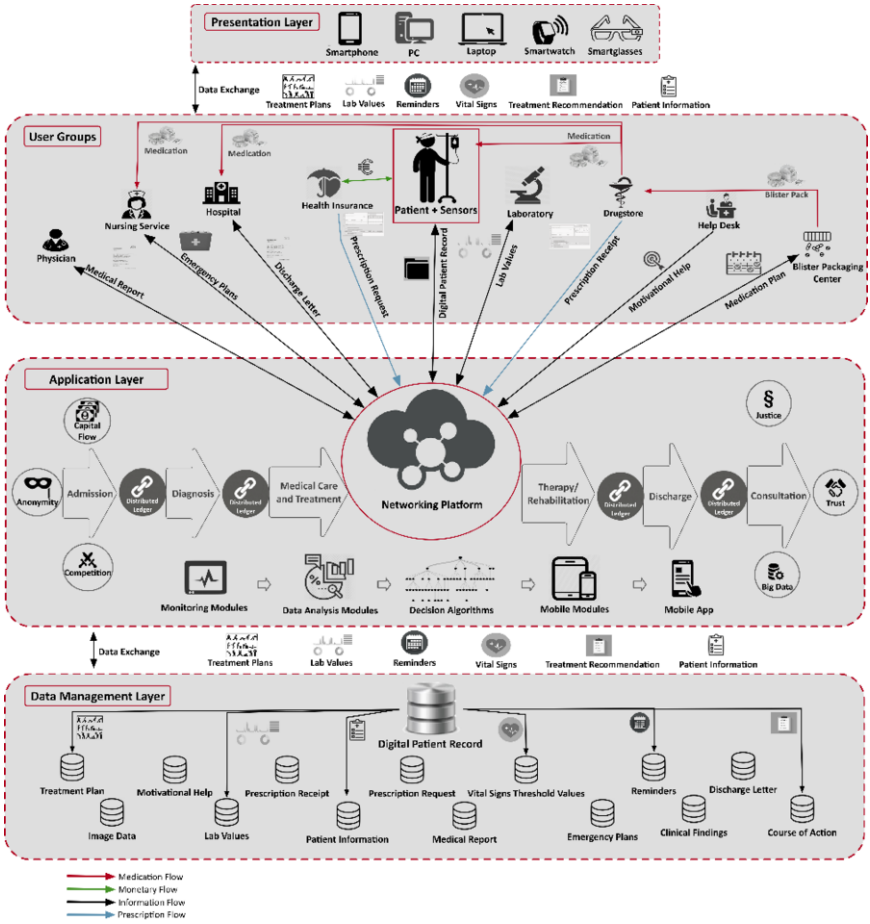


Figure 1. Telemonitoring networking platform for an integrated care concept

Another special feature is literally the distributed ledger. In the context of distributed data storage, every user in a peer-to-peer network has a complete copy of the hash functions. This eliminates the need for a central administrator and central data storage. Since it is not possible to manipulate the majority of all copies of hash functions that exist for users, the distributed data storage becomes forgery-proof through the distributed ledger. In this respect, a platform based on distributed ledger technology ultimately checks and legitimizes itself. With distributed ledger technology, there is no centralized authority with access to the data. Patients will not be forced to release data as the access rights to the digital patient file will always remain in the possession of the patients. Patients decide which health professionals can view which data and when.

4. Discussion

The paper presented the user and legal requirements for a comprehensive platform to be used by a large variety of health professionals and the patient in diabetes care. The user requirements revealed a wide spectrum of applications with the "one platform for all" as a common denominator. Although this platform is generic, diabetes is a condition that exemplifies the need of health networks, information continuity, patient-centered care and patients as owners of their data. With the new European data protection regulation, new demands emerge in particular the patient's right of data portability. Based on these requirements, a secure technology for a data backbone is needed. To this end, distributed ledger technology is proposed. The next steps for designing the platform embrace the definition of a common information model and using IT-communication standards such as HL7 FHIR and clinical terminologies such as SNOMED CT to ensure interoperability [42]. Furthermore, the distributed ledger technology needs to be specified in greater detail and exemplar use cases derived from the user requirements to be designed. We will focus on an application that show cases an approach in which the patient is the focal point and master of demanding and distributing the data and sharing decisions.

References

[1] Y. Zheng, S.H. Ley, F.B. Hu, Global aetiology and epidemiology of type 2 diabetes mellitus and its complications, Nature reviews. Endocrinology 14 (2018), 88–98.
[2] H.H. Abholz, G. Egidi, F.A. Gries, N. Haller, C. Khan, R. Landgraf et al., Nationale VersorgungsLeitlinie Therapie des Typ-2-Diabetes (Teil 2), Diabetologie und Stoffwechsel 9 (2014), 241–299.
[3] American Diabetes Association, Standards of medical care in diabetes--2010, Diabetes care 33 Suppl 1 (2010), S11-61.
[4] D. Rodbard, Continuous Glucose Monitoring: A Review of Successes, Challenges, and Opportunities, Diabetes technology & therapeutics 18 Suppl 2 (2016), S3-S13.
[5] Medtronic MiniMed, Inc., [January 23, 2018], http://www.medtronicdiabetes.com/home.
[6] R.B. Kumar, N.D. Goren, D.E. Stark, D.P. Wall, C.A. Longhurst, Automated integration of continuous glucose monitor data in the electronic health record using consumer technology, Journal of the American Medical Informatics Association JAMIA 23 (2016), 532–537.
[7] Dexcom, Inc., [January 23, 2018], https://www.dexcom.com/dexcom-g4-platinum-share.
[8] M.J. O'Brien, J.Y. Lee, M.R. Carnethon, R.T. Ackermann, M.C. Vargas, A. Hamilton et al., Detecting Dysglycemia Using the 2015 United States Preventive Services Task Force Screening Criteria: A Cohort Analysis of Community Health Center Patients, PLoS medicine 13 (2016), e1002074.
[9] J.B. Albu, N. Sohler, R. Li, X. Li, E. Young, E.W. Gregg et al., An Interrupted Time Series Analysis to Determine the Effect of an Electronic Health Record-Based Intervention on Appropriate Screening for Type 2 Diabetes in Urban Primary Care Clinics in New York City, Diabetes care 40 (2017), 1058–1064.

[10] P.J. O'Connor, J.M. Sperl-Hillen, W.A. Rush, P.E. Johnson, G.H. Amundson, S.E. Asche et al., Impact of electronic health record clinical decision support on diabetes care: A randomized trial, Annals of family medicine 9 (2011), 12–21.
[11] V. Weber, F. Bloom, S. Pierdon, C. Wood, Employing the electronic health record to improve diabetes care: A multifaceted intervention in an integrated delivery system, Journal of general internal medicine 23 (2008), 379–382.
[12] R. Benkert, P. Dennehy, J. White, A. Hamilton, C. Tanner, J.M. Pohl, Diabetes and hypertension quality measurement in four safety-net sites: Lessons learned after implementation of the same commercial electronic health record, Applied clinical informatics 5 (2014), 757–772.
[13] V.G. Vimalananda, G. Gupte, S.M. Seraj, J. Orlander, D. Berlowitz, B.G. Fincke et al., Electronic consultations (e-consults) to improve access to specialty care: A systematic review and narrative synthesis, Journal of telemedicine and telecare 21 (2015), 323–330.
[14] W. Oh, E. Kim, M.R. Castro, P.J. Caraballo, V. Kumar, M.S. Steinbach et al., Type 2 Diabetes Mellitus Trajectories and Associated Risks, Big data 4 (2016), 25–30.
[15] H.S. Glauber, N. Rishe, E. Karnieli, Introduction to personalized medicine in diabetes mellitus, Rambam Maimonides medical journal 5 (2014), e0002.
[16] I. Featherstone, J. Keen, Do integrated record systems lead to integrated services? An observational study of a multi-professional system in a diabetes service, International journal of medical informatics 81 (2012), 45–52.
[17] E.K. Christiansen, E. Skipenes, M.F. Hausken, S. Skeie, T. Østbye, M.M. Iversen, Shared Electronic Health Record Systems: Key Legal and Security Challenges, Journal of diabetes science and technology 11 (2017), 1234–1239.
[18] C. Rinner, S.K. Sauter, G. Endel, G. Heinze, S. Thurner, P. Klimek et al., Improving the informational continuity of care in diabetes mellitus treatment with a nationwide Shared EHR system: Estimates from Austrian claims data, International journal of medical informatics 92 (2016), 44–53.
[19] R.H. von Alan, S.T. March, J. Park, S. Ram, Design science in information systems research, MIS Q 28 (2004), 75–105.
[20] P. Fettke, State-of-the-Art des State-of-the-Art, Wirtsch Inform 48 (2006), 431.
[21] M.D. Myers, Qualitative research in business and management, Sage, 2013.
[22] B. Otto, H. Österle, Relevance through consortium research? findings from an expert interview study, Global Perspectives on Design Science Research (2010), 16–30.
[23] T. Ermakova, B. Fabian, R. Zarnekow, Security and privacy system requirements for adopting cloud computing in healthcare data sharing scenarios (2013),
[24] R. Warren, K. Carlisle, G. Mihala, P.A. Scuffham, Effects of telemonitoring on glycaemic control and healthcare costs in type 2 diabetes: A randomised controlled trial, J Telemed Telecare (2017), 1357633X17723943.
[25] R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: Review, opportunities and challenges, Briefings in bioinformatics (2017),
[26] M.-P. Gagnon, P. Ngangue, J. Payne-Gagnon, M. Desmartis, m-Health adoption by healthcare professionals: a systematic review, J Am Med Inform Assoc 23 (2016), 212–220.
[27] DIMDI, Sammlung Medizinprodukterecht, [January 29, 2018], http://www.dimdi.de/static/de/mpg/recht/.
[28] BGH, NJW-RR, 2014, 46 (47).
[29] Rübsamen, MedR 2015, 485 (486).
[30] Dietel/Lewalter, PharmR 2017, 53 (54).
[31] Jandt/Hohmann, KuR 2015, 694 (699).
[32] Ortner/Daubenbüchel, NJW 2016, 2918 (2919).
[33] von Schenck/Mueller-Stöfen, GWR 2017, 171 (171).
[34] C. Bauer, F. Eickmeier, M. Eckard, E-Health: Datenschutz und Datensicherheit: Herausforderungen und Lösungen im IoT-Zeitalter, Springer Fachmedien Wiesbaden, Wiesbaden, 2018.
[35] Dammann, ZD 2016, 307 (307 et seq.).
[36] von Schenck/Mueller-Stöfen, GWR 2017, 171 (174 et seq.).
[37] EuGH, 13.05.2014 - C-131/12, juris.
[38] Dammann, ZD 2016, 307 (308).
[39] LG Mosbach, 14.06.2017 - 4 O 18/17.
[40] BVerwG, PharmR 2010, 462 et seq.
[41] F. Teuteberg, S. Tönnissen, Smart Contracts, in: Das Wirtschaftsstudium, 2017, pp. 566–567.
[42] T. Benson, Using SNOMED and HL7 together, in: Principles of Health Interoperability HL7 and SNOMED, 2012, pp. 267-280.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-247

Development of a Clinical Decision Support System in Intensive Care

Lucas PFLANZL-KNIZACEK[a,b,1], Katharina BERGMOSER[a,c], Karin MATTERSDORFER[a,b], Gernot SCHILCHER[d], Christian BAUMGARTNER[c]
[a]*CBmed – Center for Biomarker Research in Medicine, Graz*
[b]*Medical University of Graz, Division of Endocrinology and Diabetology, Graz*
[c]*Graz University of Technology – Institute of Health Care Engineering with European Testing Center of Medical Devices, Graz*
[d]*Medical University of Graz – Intensive Care Unit, Department of Internal Medicine, Graz*

Abstract. Background: Intensive care is confronted with an increasing complexity and large amounts of data provided by new technological tools. One way of assisting health care professionals is providing effective clinical decision support (CDS) systems. Objectives: The aim is to develop a tailored model for the sustainable development of a clinical decision support system in intensive care. Methods: The model consists of two parts. The first part includes the interaction of the following partners: science industry and HCP. The second part comprises a three-phase process consisting of (1) the identification of clinical needs, (2) modeling and prototyping, and (3) implementation. Results: By July 2015, a government funded CDS development project started in Graz, Austria. After assigning a multi-professional and interdisciplinary team, a clinical need statement was formulated within the first six months. A prototype was developed by end of 2016 and verified using a clinical dataset. Conclusion: The developed model proofed to be feasible regarding the first two phases. Additional progress needs to made to assess the performance of the model in the implementation phase.

Keywords. Intensive Care, Clinical Decision Support, Technology Development.

1. Introduction

Modern medicine is moving towards a growing recognition of personalized and precision medicine. This reflects the emergence of a rapidly accelerating field that will leave a major imprint on the practice of medicine [1]. New technological tools have the potential to collect large amounts of digital data from different perspectives. Together with the increasing availability of molecular information [2] and the usage of advanced sensors for physiological parameter registrations, new prospects are offered to researchers and scientists in healthcare. Consequently, health care professionals (HCP) ultimately need new electronic systems processing clinical data for daily practice.

Research in intensive care has been increasing strongly since the 1980's. Figure 1 depicts the number of publications per year over time for the keywords "intensive care" and "critical care" in PubMed showing an exponential increase over time. Thus, it has become more and more difficult for clinicians to keep an overview of state-of-the-art knowledge regarding diagnosis and treatments.

[1] Corresponding Author: Lucas Pflanzl-Knizacek, Center for Biomarker Research in Medicine CBmed GmbH, Stiftingtalstrasse 5, 8010 Graz, Austria, E-Mail: lucas.pflanzl@cbmed.at

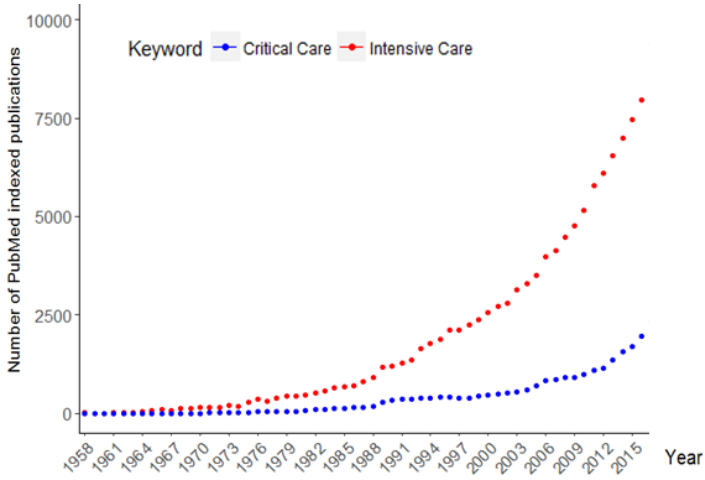


Figure 1. Yearly number of PubMed indexed publications from 1956 to 2017 for the keywords "Critical Care" and "Intensive Care" in title and abstract (TIAB search pattern).

One way to assist clinicians in their daily practice and thus improve the quality of medical care in hospitals is the use of clinical decision support systems [3]. The term "Clinical Decision Support" (CDS) as used from hereon is defined as a system "providing clinicians, patients or individuals with knowledge and person-specific or population information, intelligently filtered or presented at appropriate times, to foster better health processes, better individual patient care, and better population health" [4]. This definition refers to the consensus published along with the national roadmap on clinical decision support in 2007 in the United States.

1.1. Types of Clinical Decision Support

When approaching the different applications of CDS, it is crucial to narrow it down to the relevant types in each setting. One way of classifying decision support is to divide it into four phases [5]. Therein, CDS (1) may act as a single system, (2) be integrated into clinical systems, (3) use standards for sharing content and (4) provide service models. A more precise way is to cluster CDS systems according to their capabilities [6]. Table 1 gives an overview of the taxonomy of the systems, based on various capabilities.

Table 1. Clinical decision support taxonomy based and modified on [6].

Decision support capability	Related types
Medication dosing support	Functions for e.g. adjustment, range of dose, maximum dose or indication-based dosing.
Order facilitators	Indication-based, protocol-based or condition based ordering.
Point-of-care alerts/reminders	Drug interactions, care planning, physiological parameters, critical values
Relevant information display	Context-sensitive information retrieval, patient-specific data displays, medication cost display or context-sensitive interfaces.
Expert systems	Support functions, e.g. interpretation of sensor data and information systems, diagnostic support, treatment planning, risk assessment, prediction or interpretation.
Workflow support	Reconciliation of medication, order routing and approval, assistance in documentation.

Depending on the type of capability, CDS may range from simple methods such as dose range checking for medication, to complex ones like prognostic tools used in expert systems. The approach proposed in this paper refers to the development of expert systems. In general, CDS is able to improve the performance of HCP. A systematic review including one hundred studies showed that in 64% of the studies the practitioner performance improved by using CDS [7]. Studies included in the review assessed diagnostic, reminder, disease management and drug-dosing or prescribing systems. Regarding patient outcome, another analysis showed that a positive impact was present in 25 out of 82 studies included in systematic reviews [8].

1.2. Clinical Decision Support in Intensive Care

Treatment in intensive care units (ICU) is different from standard wards. Resources, time, workload and staffing are limited and crucial in the treatment of critically ill. Benefits for the use of CDS in intensive care have been demonstrated. One prospective controlled intervention cohort study was able to show a reduction in the incident of drug-drug interactions and related adverse events due to the use of CDS [9]. A randomized controlled trial (RCT) conducted in three different ICU across Europe demonstrated that a fully automated algorithm for tight glycemic control is safe and effective in the treatment of critically ill patients [10]. Regarding argument-based recommendations for diagnosis, a computer-interpretable guideline model for hyponatremia improved agreement with expert consensus in comparison to a paper algorithm [11]. Thus, CDS in intensive care represents a suitable option in assisting clinicians in their daily practice.

1.3. Challenges in Clinical Decision Support

The following efforts and challenges need to be considered when developing a CDS due to user and market adoption and particularly to the diversity of stakeholders involved:

- Clinical need: Early consultation of health care professionals is essential for identifying a clear clinical need [12]. Assessing the currently unmet requirements and not losing sight of them represents one of the top priorities.
- Continuous involvement of HCP: Integration starting with the beginning of the development, but also maintained longitudinally across the whole process. Including HCP in sprints or loops may lead to the reconsideration of function and application fostering sustainability.
- Interaction with existing systems: A high number of electronic systems exist in intensive care bedsides. These include hospital information and patient data management systems. In order to ensure an easy installation of the CDS system and to enable communication within a diverse technological environment, standards and state-of-the-art interfaces should be implemented.
- Availability of technology and devices: Any new device, sensor or machine, as well as any upgrade is related to monetary investments. Moreover, new systems often require additional electronic interfaces for processing data.
- Meeting the patients' need: CDS systems should also contribute to a safe and comfortable environment allowing the patients to recover. In intensive care, several physical and psychological factors represent stressors for patients [13]. Among the five greatest stressors are noise and invasive actions (such as tubes) [14]. New systems should aim at improving patient quality by using minimally

invasive methods and intelligent alarms. Alarms in general may lead to alarm fatigue by the users, which in turn affects patient safety [1516]. Prompts and alarms need to be tailored to fit patient and user needs [17].

These first discussion points do not provide an entire overview regarding challenges in CDS (see [18] for a detailed analysis).

1.4. Aims

The main aim of our approach was to propose a tailored model for the sustainable development of a CDS system in intensive care. In contrast to normal wards, the field of intensive care represents a high-reliability environment [19], bringing along a whole set of crucial requirements. Secondly, the model has to relate to the various challenges associated with CDS described under 1.3. In summary, following the model shall foster the development of a viable system, independent of the addressed type of CDS.

2. Methods

We consider two aspects vital for an integrated development approach. First, a pool of key partners need to be included by taking into account the multi-professional and interdisciplinary environment in intensive care, representing the cornerstones of the model. The second aspect of importance represents the process of developing the system. This may include different phases or iterative loops, from ideation until market entry. Regarding medical technologies, a step-by-step model is provided by the Stanford Biodesign [20], consisting of three distinct phases: "identify, invent and implement". The process model we propose has familiarities, whereas it focuses more on the developmental approach, not taking into account the economic and legal perspective.

2.1. Interaction of partners

Our model includes scientific partners, health care professionals and industrial partners. Figure 2 depicts their interactions, with a common interaction area in the center of the model. All three different fields are to cooperate within the development of the CDS. HCP provide insight into daily practice. This is necessary to hit the clinical need when starting to identify the field of action. Their continuous involvement throughout the whole process guarantees practical gains for the clinical applicability and reliefs common problems. The scientific partner develops the models required for the expert system, specifically designed and tailored to the needs of the HCP. Finally, the industrial partner provides the framework for technological feasibility and implementation.

(1) Interaction between the scientific and the industrial partner. Focusing on evidence-based fundamentals, the scientific partner provides knowledge required by the industrial partner for technically developing the CDS. Relevant data needed for developing models can be provided by conducting clinical trials.

(2) Interaction between the scientific partner and health care professionals. The dialogue between the scientific partner and HCP is particularly important for proofing the applicability of concepts and models. HCP can pinpoint problems in clinical practice and feedback on the effectiveness and efficiency of a conceptual model, thereby fitting it to their needs. Moreover, the scientific partner can be invited to join providers in their daily work by carrying out walkthroughs to identify problems.

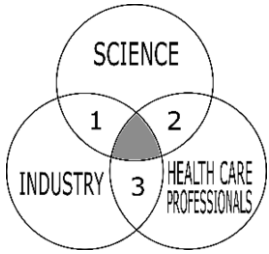


Figure 2. Interaction of partners for CDS development. (1), (2) and (3) showing overlaps for cooperation.

(3) Interaction between health care professionals and industry. Along the progress of developing the technology for the CDS, HCP can assist the industrial partner in regularly testing usability and design. By integrating HCP in a very early phase, time and effort spent on otherwise later tests will lead to a quicker development.

2.2. Process model

The process used for CDS development consists of three distinct phases (see Figure 3). At first, identifying the clinical needs is of fundamental importance for the whole process as it already relates directly to daily clinical practice. Secondly, the model for the CDS expert system is set up, accompanied by a simultaneous and quick prototyping, integrating HCP in feedback loops. In the third phase, the implementation takes place.

Throughout every phase, the plan-do-study-act (PDSA) cycle [21] is used. The hypothesis of each phase is tested and adapted due to the results obtained using the PDSA cycle. This allows a profound development along each phase, which may reject, restate or fine-tune the underlying hypothesis used for CDS development.

(1) Identification of clinical needs. The scientific partner is cooperating on a close basis with HCP. Underlying problems in the clinical setting are analyzed and structured, e.g. by qualitative interviews or walkthroughs. Additionally, current scientific literature and latest publications are screened for understanding the context of the problem field. A justification of the identified problem has to be formulated in order to have a firm basis for further investigation. Planning clinical trials at the end of this phase may assist if there is a need for more or additional data or information, respectively.

(2) Modeling and prototyping. After having formulated the clinical need, a conceptual CDS model is set up. The scientific and industrial partner work together to develop a first prototype, without functionality towards the patient, but providing a tool to assess its applicability together with HCP. Advantages of such a non-functional prototype are the early availability for testing and receiving feedback of HCP, saving costs and time instead of long development runs and a quick adaption and integration of changes based on expert input. Clinical trials provide data and information for the

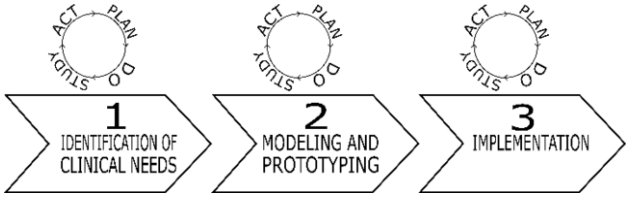


Figure 3. Process model for the CDS development.

development of the underlying model, taking place in parallel to usability tests of the non-functional prototype. Data sources may include the hospital information system, electronic health record or patient data management systems.

(3) Implementation. The main focus lies on directly translating the prototype including the model into a functional product. Existing sensors and devices are considered to reduce invasiveness and additional connections towards the patient. Together with the scientific partner and HCP, the industrial partner is able to conduct clinical trials for means of evaluation. Planning for such trials can start in parallel to the implementation, increasing efficiency in comparison to traditional approaches.

3. Results

By July 2015 a CDS project was started in the framework of the government funded COMET-K1 Centre, Center for Biomarker Research in Medicine (CBmed) in Graz, Austria. With CBmed as lead, B.Braun Melsungen AG as industrial, Medical University of Graz (MUG) and Graz University of Technology as scientific partners have joined the project. A multi-professional and interdisciplinary team consisting of biomedical engineers (1.75 full time equivalents (FTE)), intensivists (available at MUG), nurses (0.125 FTE and available at MUG), molecular biologists (0.75 FTE) and clinical trial experts (0.5 FTE) has been assigned to the project. The FTE refer to the employment at CBmed. Personnel of industrial and scientific partners join based on pre-specified in-kind contributions. The core team holds representatives of each partner (one per partner), setting the stage for close cooperation within every phase. All team members work within the MUG Campus. Arranging regular meetings and jour fixes over short time spans, every partner involved is up to date about the progress of each other. The core project team jour fixe takes place weekly, whereas coordination with the industrial partner takes place in a bi-weekly meeting schedule. Medical advisors such as intensivists or nurses are kept up to date on a monthly basis or are invited to the core project team jour fixe on demand.

The identification of clinical needs was carried out within the first six months, mainly by multi-site quality interviews conducted with intensivists and nurses. A total number of 12 HCP were interviewed in two different Austrian hospitals. Using these results, a robust justification report was created wherein the clinical need statement was related to the specified area of research. Modeling and prototyping started by beginning of 2016. A prospective monocentric observational clinical trial (Clinibil, clinicaltrials.gov identifier NCT02914782) was planned and started in September 2016 at an intensive care unit of MUG to provide extensive clinical data needed for the model. The trial includes electronic data about patient demographics, individual data of medical sensors during the ICU stay (i.e. heart rate, blood pressure, invasive circulation monitoring parameters, ventilation parameters, lab test results for serum and urine including electrolytes) and information about the medical treatment (i.e. administration of drugs and fluids, invasive interventions). Additionally, biosamples of the patients were acquired and aliquots stored at BioBank Graz for later in-depth analysis. By end of 2016, a first non-functional prototype was available. In 2017, the prototype was tested in further multi-site qualitative interviews with intensivists and nurses of the first interview round. In parallel, the development of a mathematical model for the analysis of the clinical course of selected parameters including fluid and electrolyte management started. Data acquired in Clinibil for 52 patients was used for early verification of the developed model

by end of 2017. Verification was carried out by electronically comparing the predicted results of the model to the data of the clinical course of each patient included in Clinibil at specified time points and timeframes. Next steps for enhancing the model and validating its applicability by a second trial including more than 2000 patients are planned to finish the second phase by end of 2018. This includes combining the user interface of the non-functional prototype together with the validated model. The project can then transit to the implementation phase for the finalization of a new medical device including necessary evaluations in situ.

4. Discussion

The increased amount of scientific work published in the field of intensive care together with the growing share of electronic systems and devices makes it necessary to adapt the way CDS systems are developed. Traditional ways of research and development (R&D), starting in R&D divisions without an early involvement of HCP, may hamper the development of devices and systems that tackle problems of daily practice and take into account the clinical needs. The approach for CDS development proposed in this paper tries to consider these needs by integrating a multi-professional and interdisciplinary partnership model into a three-phase process. The early identification of clinical needs makes it possible to avoid unnecessary development iterations. Moreover, continuous feedback and evaluation by HCP provide the possibility to take into account the applicability of the CDS in comparison to already used systems, tailoring it specifically to specified problems of clinical practice.

The results show that the approach for CDS development is feasible. Regarding the time span of the various phases, it has to be highlighted that formulating a clinical needs statement is possible within a rather short time span as of six months. Therefore, modeling and prototyping can start early. Moreover, the availability of clinical data from the conducted trials provide tailored data for development of the CDS model. However, as the project has not yet begun with the implementation phase, no statement about efficacy in terms of a final CDS system is possible. Saving of time and resources by constantly involving HCP can only be evaluated after project end. Concerning the mode of cooperation between the partners, mutual understanding has been set up between the various and different professions involved. This is supported by the rather high frequency of meetings and jour fixes, thereby constantly keeping track with each project partner, possibly proofing beneficial in respect to the sustainability of the new system.

In clinical practice, CDS will definitely become more important within the next decade. Electronic systems will not only support, but also routinely provide services especially in settings with high workload, such as intensive care. Therefore, these systems have to fulfill the needs of HCP and patients. The described approach for CDS development is one way of integrating different partners to accomplish such goals.

Acknowledgements

"CBmed" is cooperating together with B.Braun Melsungen AG for this research project. Work done in "CBmed" was funded by the Austrian Federal Government within the COMET K1 Centre Program, Land Steiermark and Land Wien.

References

[1] J.L. Jameson, D.L. Longo, Precision Medicine - Personalized, Problematic, and Promising, The New England Journal of Medicine 372(23) (2015), 2229-2234.
[2] L.L. Bu, K. Yang, W.X. Xiong, F.T. Liu, B. Anderson, Y. Wang, J. Wang, Toward precision medicine in Parkinson's disease, Annals of Translational Medicine 4(2) (2016), 26.
[3] M. Khalifa, Clinical Decision Support: Strategies for Success, Procedia Computer Science 37(2014), 422-427.
[4] J.A. Osheroff, J.M. Teich, B. Middleton, E.B. Steen, A. Wright, D.E. Detmer, A roadmap for national action on clinical decision support, Journal of the American Medical Informatics Association 14(2) (2007), 141–145.
[5] A. Wright, D.F. Sittig, A four-phase model of the evolution of clinical decision support architectures, International Journal of Medical Informatics 77(10) (2008), 641-649.
[6] A. Wright, D.F. Sittig, J.S. Ash, J. Feblowitz, S. Meltzer, C. McMullen, K. Guappone, J. Carpenter, J. Richardson, L. Simonaitis, R.S. Evans, W.P. Nichol, B. Middleton, Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems, Journal of the American Medical Informatics Association 18(2011), 232-242.
[7] A.X. Garg, N.K.J. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J. Beyene, J. Sam, R.B. Haynes, Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review, Journal of the American Medical Association 293(10) (2005), 1223-1238.
[8] M.W.M. Jaspers, M. Smeulers, H. Vermeulen, L.W. Peute, Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings, Journal of the American Medical Informatics Association 18(3) (2011), 327-334.
[9] T. Bertsche, J. Pfaff, P. Schiller, J. Kaltschmidt, M.G. Pruszydlo, W. Stremmel, I. Walter-Sack, W. Haefeli, J. Encke, Prevention of adverse drug reactions in intensive care patients by personal intervention based on an electronic clinical decision support system, Intensive care medicine 36 (2010), 665-672.
[10] J. Plank, J. Blaha, J. Cordingley, M.E. Wilinska, L.J. Chassin, C. Morgan, S. Squire, M. Haluzik, J. Kremen, S. Svacina, W. Toller, A. Plasnik, M. Ellmerer, R. Hovorka, T.R. Pieber, Multicentric, randomized, controlled trial to evaluate blood glucose control by the model predictive control algorithm versus routine glucose management protocols in intensive care unit patients, Diabetes Care 29(2) (2006), 271-276.
[11] A.G. González-Ferrer, M.A. Valcárcel, M. Cuesta, J. Cháfer, I. Runkle, Development of a computer-interpretable clinical guideline model for decision support in the differential diagnosis of hyponatremia, International Journal of Medical Informatics 103() (2017), 55-64.
[12] J.L. Martin, D.J. Clark, S.P. Morgan, J.A. Crowe, E. Murphy, A user-centred approach to requirements elicitation in medical device development: a case study from an industry perspective, Applied Ergonomics 43(1) (2012), 184-190.
[13] P. Soehren, Stressors perceived by cardiac surgical patients in the intensive care unit, American Journal of Critical Care 4(1) (1995), 71-76.
[14] M.A.F.P. Novaes, A. Aronovich, M.B. Ferraz, E. Knobel, Stressors in ICU: patients' evaluation, Intensive Care Medicine 23(12) (1997), 1282-1285.
[15] K.J. Ruskin, D. Hueske-Kraus, Alarm fatigue: impacts on patient safety, Current Opinion in Anesthesiology 28(6) (2015), 685-690.
[16] Joint Commission, Medical device alarm safety in hospitals, Sentinel Event Alert/Joint Commission on Accreditation of Healthcare Organizations 50 (2013), 1-3.
[17] B.J. Drew, P. Harris, J.K. Zègre-Hemsey, T. Mammone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, X. Hu, Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients, PLOS ONE 9(10) (2014), e110274.
[18] D.F. Sittig, A. Wright, J.A. Osheroff, B. Middleton, J.M. Teich, J.S. Ash, E. Campbell, D.W. Bates, Grand challenges in clinical decision support, Journal of Biomedical Informatics 41 (2008), 387-392.
[19] K.H. Roberts, Some Characteristics of High-Reliability Organizations, Organization Science 1 (1990) 160-177.
[20] P.G. Yock, S. Zenios, J. Makower, T.J. Brinton, U.N. Kumar, F.T.J. Watkins, L. Denend, T.M. Krummel, C.Q. Kurihara, Biodesign - The process of innovating medical technologies, Cambridge University Press, Cambridge United Kingdom, 2015.
[21] W.E. Deming, Out of the Crisis, MIT Press, Cambridge Massachusetts, 1986. p. 88.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-255

Development and Evaluation of Cognitive Games to Promote Health and Wellbeing in Elderly People with Mild Cognitive Impairment

Mark SCASE[a,1], Karl KREINER[b] and Antonio ASCOLESE[c]
[a]*Division of Psychology, De Montfort University, Leicester, UK*
[b]*Center for Health & Bioresources, Austrian Institute of Technology, Graz*
[c]*imaginary srl,Milano, Italy*

Abstract. Background: In Europe the number of elderly people is increasing. This population growth has resulted in higher healthcare costs. The purpose of this project was to try to promote active ageing in people aged 65-80 with mild cognitive impairment through cognitive games delivered via a tablet computer. Objectives: Age-appropriate cognitive games were developed targeting different aspects of cognition and then experiences of elderly people using these games were evaluated. Methods: The design of games was developed through iterative user-centered design focus groups with elderly people as participants. The experiences of participants playing the games over a 47 day period were explored through semi-structured interviews. Results: Four games were developed that addressed a range of cognitive functions such as perception, attention, memory, language, comprehension and executive function. The participants were able to play these games without external intervention over an extended period and reported positively on their experiences. Conclusion: Cognitive games can be used successfully by people with mild cognitive impairment to promote active ageing.

Keywords. Psychology, cognition, aging, mobile application, eHealth.

1. Introduction

The World Health Organization has suggested that the proportion of the world's population over 60 years old will nearly double from 12% to 22% between 2015 and 2050 and the number of people aged 60 years and older will outnumber children younger than 5 years [1]. The ageing population of Europe is increasing as well [2]. This demographic shift will pose challenges to the health and social systems in many countries. As individuals age so their healthcare costs increase [3]. The health of elderly people is determined by many components (e.g. mental, physical, interpersonal relationships etc.) and to improve their overall lives active ageing is encouraged [4]. This paper describes the development and implementation of cognitive games for a project designed to enhance the health and wellbeing of elderly people aged 65-80 year with mild cognitive impairment. The DOREMI project [5] (**D**ecrease of c**O**gnitive decline, malnut**R**ition and sed**E**ntariness by elderly empowerment in lifestyle **M**anagement and social **I**nclusion) combined scientists from different disciplines to produce activities targeting

[1] Corresponding Author: Mark Scase, Division of Psychology, De Montfort University, The Gateway, Leicester, LE1 9BH, UK, E-Mail: mscase@dmu.ac.uk

health and wellbeing in elderly people. The increased costs of health and social care in elderly people can arise from poor nutrition, sedentariness, cognitive decline and social isolation [6]. A lifestyle intervention was delivered by DOREMI on a tablet computer with the aim to encourage active ageing by increasing socialization and physical activity, promoting cognitive function and improving diet [7]. The burden on healthcare systems could then be reduced by encouraging participants to be autonomous with the program delivered by a tablet computer [8]. Participants can be motivated to continue with the program via gamification [9-11] – the use of age-appropriate design elements to increase activity and participation.

This paper describes one element of the DOREMI project [12]: the design, development and implementation of cognitive games for a specific group of elderly people with mild cognitive impairment. Health related games have been produced previously [13] but this project was novel in that individuals with profiles the same as the intended users were involved in the design and development. The theoretical approach of this project was one rooted in that of positive psychology and psychological wellbeing [14] which is to improve individual's lives. In order for people to engage in the task the games had to be designed in a way that was relevant to the participant with an appropriate level of difficulty. Successful implementation would result in participants continuing with the task because of high self-efficacy [15]. Participants would reach a state called "Flow", a term from wellbeing theory [16] where they would be self-motivated and make positive health choices [17].

There has been considerable debate on the effectiveness of computerized cognitive games on producing any change in cognitive function. For example, one systematic review of research with unimpaired elderly adults found that the use of computerized training was more effective than that of traditional paper-and-pencil cognitive training [18]. Other systematic reviews in cognitively healthy adults concluded that the experimental design was important in determining the outcome [19-20]. One of the most recent and extensive reviews on "brain-training" concluded that interventions did improve performance but there was not generalization across cognitive domains [21]. There have been some systematic reviews of training with people either at risk of dementia or elderly adults with mild cognitive impairment [22-23]. These studies suggested there might be moderate positive effects. Overall, many of these papers recommended there should be further research.

There were two main aims to this study: first to design age-appropriate cognitive games through a user-centered design that could be delivered via a tablet computer; and second to explore the experiences of elderly people with of this technology solution after a 47 day period. The preferences of individuals were first investigated through a series of iterative user-centered focus groups. Qualitative data were collected on the experiences of the participants through individual interviews. Our research questions were as follows:- 1. What components of a cognitive game are preferred by elderly adults and how can these people be used to help in the design of cognitive games? 2. What are the experiences of elderly people playing cognitive games on a tablet computer over a 47 day period?

2. Methods

2.1. Development of cognitive games

An iterative user-centered design process was used in three focus groups to develop the cognitive games. Participants had normal cognition or mild cognitive impairment as assessed by the mini mental state examination (MMSE) [24] or the Montreal Cognitive Assessment (MoCA) [25]. In the first focus group there were nine people (5 male; mean age=77.0, SD=7.47; mean MMSE=29.3, SD=1.00). This group found out about the exposure to technology of individuals and ideas for cognitive games. In the second focus group there were five people (1 male; mean age=74.6, SD=5.46; mean MoCA=22.8, SD=1.64). This group tested initial games and received feedback on user experience. There were four people in the third focus group (1 male; mean age=78.5, SD=1.91; mean MoCA=22.0, SD=2.45). This group tested games that had been modified following suggestions from the first group and obtained more participant input. The groups were facilitated by two psychologists with each session lasting 50 to 70 minutes. The recordings were transcribed and analyzed thematically.

2.2. Testing of cognitive games

Following the final development of the cognitive games the experiences of older people playing these games was explored over an extended period. Participants aged 65-80 years old with mild cognitive impairment were recruited to be part of the study. All participants were living alone. There were 25 participants (3 male; mean age=75.0, SD 4.28; mean MoCA 24.2, SD=1.71).

Participants had a training period over 16 days when they received advice on how to use the tablet computer and engagement with the games was recorded remotely. There then followed a period of 47 days during which these participants followed the DOREMI protocol which included playing cognitive games.

2.3. Post-Intervention Participant Interviews

The experiences of participants in the project were explored through individual interviews lasting an average of 35 minutes. The interview schedule was designed to elicit specific and detailed responses. Interviews were transcribed verbatim and analyzed using thematic analysis which allowed for participant common themes to be identified. An inductive approach used so that the identified themes were data driven. Thematic analysis steps were followed as previously detailed [26] involving the reading and re-reading of the transcripts, followed by the generation of codes. Participant responses were then classified according to these codes, after which codes were collapsed to form themes.

Figure 1. "Find it" cognitive game. In this example participants are asked to touch all light blue rectangles.

3. Results

3.1. Cognitive games development

Six of the nine participants in the first focus group had some computer literacy with participants reporting using computers for online banking, shopping, social networking and Skype. Participants liked playing games such as puzzles, card games and trivia games and reported they did this to maintain their cognitive health. Three participants had manual dexterity problems and requested a design that was age-appropriate. Computer technology was viewed by the group as a useful way of passing the time when unable to complete normal activities due to bad health. Additionally, participants felt that games could offer some social interaction for elderly people who did not often leave the house.

Figure 2. "Match it" cognitive game. Participants are shown cards which turn over and then asked to remember the spatial position of pairs of tiles.

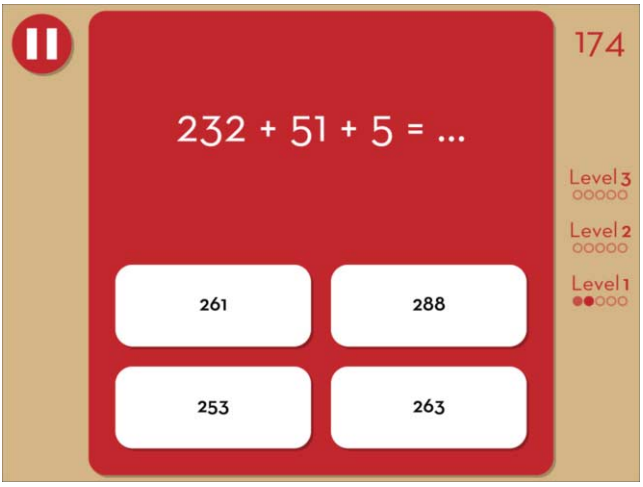


Figure 3. "Solve it" cognitive game. Participants are asked to indicate the solution to a mathematics problem by pressing on the correct answer.

The second focus group introduced participants to a prototype of a game where they were asked to find and press on objects presented on the screen. Participants reported they were able to follow the instructions well and could operate the tablet computer. Feedback was received on color selection with participants preferring bright colors.

In the third focus group two further games were introduced: one was a memory game where participants were shown pairs of cards on the screen which then turned over. The task of the participants was to remember the location of cards and select them in pairs. The other was a game on mathematical problems where answers were presented as multiple choice. Participants found the mental arithmetic challenging but liked this game. They commented that the symbol they would use for divide would be "÷" rather than "/" which is more commonly used now.

Following participant feedback four games were finalized: Find it—where participants looked for and touched screen elements as fast as possible with this game assessing language, memory and executive function (see Figure 1); Match it—a card pairing game where participants selected pairs of objects with this game designed to assess memory, attention and spatial layout (see Figure 2); Solve it—a mathematical game where participants solved problems by multiple choice which tested calculation (see Figure 3); and Complete it—where participants filled in the missing photograph pieces by dragging, dropping and if necessary rotating elements with this game testing spatial attention and memory (see Figure 4).

3.2. Evaluation of the cognitive games

Participants successfully completed the 47-day period over which they played the cognitive games by themselves and experienced the four different games.

Thematic analysis of the participant interviews after the period of playing games revealed a number of themes. Participants spoke positively about the cognitive games *"Ooh, I like them!"* (participant E01) and the overall experience *"I'm glad I had the experience"* (participant E01) and *"It's been good, I've enjoyed them"* (participant E06). There were comments that the games had been beneficial *"Well, I think it's done me*

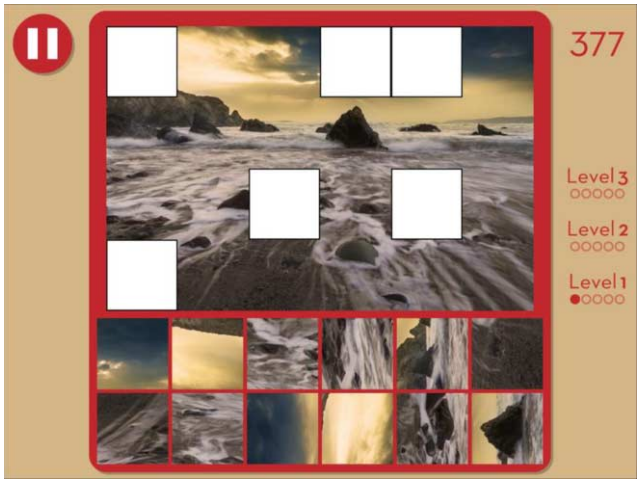


Figure 4. "Complete it" cognitive game. Participants are asked to complete a picture on the top of the screen by dragging across (and rotating if necessary) the missing parts from the bottom.

good" (participant E01), "*it broadens your mind*" (participant E02) and *"You're never too old to learn"* (participant E05). In terms of the different games "Find it" was liked "*I enjoyed that the best*" (participant E02) along with "Solve it" "*the maths were good*" (participant E02) "*the maths games was good*" (participant E05) in this case because of the link to a former career. The game "Match it" received good reports *"That one I quite liked"* (participant E08). On the other hand, "Complete it" was liked less "*I found it boring*" (participant E02).

4. Discussion

Through an iterative user-centered design a set of four cognitive games were devised for elderly people with mild cognitive impairment. This approach has been applied previously for health, education and for elderly people [27-28]. The usability of software on tablet computers with elderly people needs to be considered carefully because of changes in motor performance [29] and vision [30] with age. Therefore, controls need to be of an appropriate size with high contrast elements. Furthermore, the language used in instructions on games needs to be age-appropriate.

Qualitative results from the intervention suggested that participants felt positive about the cognitive games and there were reports that some participants felt the games had benefitted them. Cognitive games have been used preciously in studies attempting to enhance cognition in elderly people [18-23]. There have been mixed reviews of the effectiveness of games to change cognition. For example, one large systematic review [21] suggested that although games can improve performance on the trained task there is less evidence that interventions improve performance on closely related tasks. For any brain training task to be effective there needs to be generalization. Future developments could include the use of games similar to those that have been devised in this study and giving them to elderly people over an extended period of time in a controlled study (such as 6 months or more) and monitoring whether they produce any long lasting general cognitive changes.

Psychological wellbeing is a complex construct, particularly for elderly people and contains several components that interact with each other [31]. During the initial focus groups, the participants said that an important factor for maintaining health and wellbeing during old age was to keep their brains active along with regular exercise, social interaction, and a healthy diet. Certainly, the confidence that participants had with using technology improved during the time they used the games and this confidence in technology together with the experience of playing the cognitive games can enhance overall self-esteem [32] and promote active ageing.

Acknowledgments

This study is part of the EU Framework 7 DOREMI project, Grant agreement - 611650 http://www.doremi-fp7.eu/

5. References

[1] World Health Organisation, Ageing and health, http://www.who.int/mediacentre/factsheets/fs404/en/, last access 3.2.2018.
[2] European Commission, Population ageing in Europe: Facts, implications and policies. Brussels, Belgium, 2014, https://ec.europa.eu/research/social-sciences/pdf/policy_reviews/kina26426enc.pdf, last access: 3.2.2018.
[3] D.H. Yamamoto, Health Care Costs – From Birth to Death. Society of Actuaries, 2013, http://www.healthcostinstitute.org/files/Age-Curve-Study_0.pdf, last access: 3.2.2018.
[4] World Health Organization, Active Ageing: A Policy Framework, 2002.
[5] DOREMI - Decrease of cOgnitive decline, malnutRition and sedEntariness by elderly empowerment in lifestyle Management and social Inclusion. EU Seventh Framework Programme. Grant agreement no: 611650 (http://www.doremi-fp7.eu/)
[6] European Commission, Commission Communication: the Demographic Future of Europe – From Challenge to Opportunity. European Commission, Brussels, 2006.
[7] World Health Organization, Active Ageing: A Policy Framework, 2002.
[8] R. Beech, D. Roberts, Research Briefing 28: Assistive Technology and Older People. Social Care Institute for Excellence, 2008
[9] S. Deterding, D. Dixon, R Khaled, L. Nacke, From Game Design Elements to Gamefulness: Defining Gamification. Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (2011) 9-15.
[10] B. Burke, How Gamification Motivates People to Do Extraordinary Things, Bibliomotion Inc, Brookline, 2014.
[11] M. Scase, B. Marandure, J. Hancox, K. Kreiner, S. Hanke, J. Kropf, Development of and adherence to a computer-based gamified environment designed to promote health and wellbeing in older people with mild cognitive impairment, Studies in Health Technology and Informatics 236, (2017) 348-355.
[12] F. Palumbo, D. La Rosa, E. Ferro, D. Bacciu, C. Gallicchio, A. Micheli, S. Chessa, F Vozzi, O. Parodi, Reliability and human factors in Ambient Assisted Living environments. The DOREMI case study, Journal of Reliable Intelligent Environments 3(3) (2017), 139–157.
[13] D.R. Michael, S.L. Chen, Serious Games: Games that Educate, Train, and Inform. Muska & Lipman/Premier-Trade, 2005.
[14] E.L Deci, R.M. Ryan, Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press, New York, 1985.
[15] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychological Review, 84 (1977) 191-215.
[16] M. Csikszentmihalyi, Flow: The Psychology of Optimal Experience. Harper & Row, New York, 1990.
[17] M. Conner, P. Norman, Predicting Health Behaviour, McGraw-Hill Education, Berkshire, 2005.
[18] A.M. Kueider, J.M. Parisi, A.L. Gross, G.W. Rebok, Computerized Cognitive Training with Older Adults: A Systematic Review. PLoS ONE, 7(7) (2012), e40588. https://doi.org/10.1371/journal.pone.0040588

[19] A. Lampit, M. Valenzuela, N.J. Gates, Computerized Cognitive Training Is Beneficial for Older Adults, Journal of the American Geriatrics Society, 63(12) (2015), 2610-2612.
[20] A. Lampit, H. Hallock, M. Valenzuela, Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers, PLoS Med, 11(11) (2014), e1001756. https://doi.org/10.1371/journal.pmed.1001756
[21] D.J. Simons, W.R. Boot, N. Charness, S.E. Gathercole, C.F. Chabris, D.Z. Hambrick, E.A.L. Stine-Morrow, Do "Brain-Training" Programs Work? Psychological Science in the Public Interest, 17(3) (2016), 103-186.
[22] N.T. Hill, L. Mowszowski, S.L. Naismith, V.L. Chadwick, M. Valenzuela, A. Lampit, Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis, The American Journal of Psychiatry 174(4) (2017), 329-340.
[23] N.J. Gates, P.S. Sachdev, M.A. Fiatarone Singh, M. Valenzuela, Cognitive and memory training in adults at risk of dementia: a systematic review, BMC Geriatrics, 11 (2011), 55. https://doi.org/10.1186/1471-2318-11-55
[24] M.F. Folstein, S.E. Folstein, P.R. McHugh, Mini-mental State: A Practical Method for Grading the Cognitive State of Patients for the Clinician, Journal of Psychiatric Research, 12 (1975) 189-198.
[25] Z.S. Nasreddine, N.A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J.L.C. Cummings, H. Chertkow, The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. Journal of the American Geriatric Society, 53, (2005) 695–699.
[26] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. 4 Qualitative Research in Psychology, Vol. 3, No. 2, pp. 77-101.
[27] B. Lange, C. Chang, E. Suma, B. Newman, A.S. Rizzo, M. Bolas, Development and Evaluation of Low Cost Game-Based Balance Rehabilitation Tool Using the Microsoft Kinect Sensor, Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE (2011) 1831-1834.
[28] R. Profitt, B. Lange, User Centred Design and Development of a Game for Exercise in Older Adults. International Conference on Technology, Knowledge and Society (2012).
[29] K.L. Francis, P.G. MacRae, W.W. Spirduso, T. Eakin, Age and practice effects on inter-manual performance asymmetry, Frontiers in Psychology, 5 (2015), 1585. https://doi.org/10.3389/fpsyg.2014.01585
[30] N.S. Gittings, J.L. Fozard, Age Related Changes in Visual Acuity, Experimental Gerontology, 21 (1986) 423-433.
[31] S. Jivrai, J. Nazroo, B. Vanhoutte, T. Chandola, Aging and Subjective Well-Being in Later Life, The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 69(6) (2014), 930-941.
[32] D.J. Miller, D.P. Roberson, Using a games console in the primary classroom: Effects of 'Brain Training' programme on computation and self-esteem, British Journal of Educational Technology, 41(2) (2010), 242-255.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-263

Estimation of the Accuracy of Prognostic Scores for the Treatment of Children with Severe Trauma in a Specialized Trauma Hospital

Irina V. VASILYEVA[a,1], Sergey L. SHVIREV[b], Tatyana V. ZARUBINA[c], Olga V. KARASEVA [d], Alexander S. SAMOYLOV [a], Yuri D. UDALOV[a]
[a] *State Research Center - Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia*
[b] *Federal Research Institute for Health Organization and Informatics of Ministry of Health of the Russian Federation, Moscow, Russia*
[c] *The Russian State Medical University, Moscow, Russia*
[d] *Clinical and Research Institute of Urgent Pediatric Surgery and Trauma, Moscow, Russia*

Abstract. The aim of the present work was to study the validity and prognostic accuracy of scores for assessing the severity of the condition in children with severe trauma, located in the Department of Anesthesiology and Resuscitation in the Clinical and Research Institute of Urgent Pediatric Surgery and Trauma. The prospective study was conducted using clinical and physiological data collected at the admission and during the first 24 hours of hospitalization from 474 patients. The validity and prognostic accuracy of prognostic scores were assessed by determining their discrimination and calibration ability. A comparison of the discriminatory ability of scores was carried out by comparing the areas under the ROC curves with the z-criterion. Four prognostic scores were included into the study: PRISM, APACHE II, ISS-RTS-TRISS, which were used for calculating the severity of injury and for prognosis of death. Score PTS was used for evaluating the severity index only. Our results indicate that only score ISS-RTS-TRISS may be useful in practice (has excellent discrimination ability and significant calibration ability). The other lack either discrimination ability (PRISM) or calibration ability (PTS, APACHE II). The result of the study has shown that only one of the four prognostic scores, ISS-RTS-TRISS, can be successfully used in everyday practice in the department of anesthesiology and resuscitation in the specialized hospital of children's traumatology to assess the severity of the condition, with the possibility of predicting the likelihood of a lethal outcome.

Keywords: prognostic score, validity of prognostic scores, prognostic accuracy of score, ISS-RTS-TRISS, PRISM, APACHE II, PTS

1. Introduction

Trauma is the leading cause of death and disability in childhood [1]. The consequences of trauma are expressed by millions of hospitalized patients and appeals to emergency

[1] Corresponding author: Irina V. VASILYEVA, State Research Center - Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russian Federation. Email: iv001yt@gmail.com

departments. It is noted that the frequency of deaths is directly dependent on the quality of first aid at the pre-hospital and the admission department [2, 3]. At the same time, assessing the severity of the condition and severity of the injury affects the choice of tactics, the quality and effectiveness of the medical care provided. In modern medicine, the process of assessing the severity of the patient's condition and the severity of the injury is based not only on a qualitative description of the injury, but also on a quantitative assessment [4, 5]. The prognostic scores, which are mathematical models that include basic demographic, clinical, and laboratory data are an example of the quantitative assessment of the severity state [6]. For patients with injuries a large set of forecasting tools exists at this moment. The differ in terms of their intended purpose, some general and some specialized. Such promising scores are ISS-RTS-TRISS (Baker and colleagues, 1974, Boyd and colleagues, 1987) [6, 7], APACHE II (Knaus et all 1985) [8] for adults, PRISM (M. Polluck and colleagues,1988) [9] and PTS (Joseph J. Tepas and colleagues, 1987) [10] for children; the first three scores allow to define the probability of death outcome in patients.

Despite the extensive use of prognostic scores in clinical practice, the authors note the existence of open questions in this area like a dependence of scales validity on the patient's age, residence's territory, nosologies and other parameters [11, 12, 13].

The aim of the present study is to study of the validity and prognostic accuracy of scores for assessing the severity of the condition in children with severe trauma, located in the department of anesthesiology and resuscitation in the Clinical and Research Institute of Urgent Pediatric Surgery and Trauma. To achieve this aim we set following tasks: to investigate the calibrating and discriminatory abilities of the ISS-RTS-TRISS, PTS, APACHE II and PRISM prognostic scores according to the data of children with injuries treated in the Department of Anesthesiology and Resuscitation of a specialized hospital; to compare the prognostic accuracy of scores in children with injuries; to determine the points of optimal separation of patients into groups with likelihood and lethal outcomes; to assess the possibility of using prognostic scores in groups of children with different causes of injury and age.

2. Material and Methods

The prospective study was performed on basis of the Departments Anesthesiology and Resuscitation in the Clinical and Research Institute of Urgent Pediatric Surgery and Trauma and the Department of Medical Cybernetics of The Russian State Medical University. The data collected from 474 patients (mean age 9.4 ± 5.1, m/f were 303/171 patients) with trauma from the age of 1 month to 18 years old who had been admitted to the Department of Anesthesiology and Resuscitation between January 2008 and January 2012. 40 (9.6 %) patients died. Criteria for including patients into the study were the following: heavy traumatic damage, age till 18, availability of all necessary data for calculating results of prognostic scores, staying in the anesthesiology and resuscitation department not less than 24 hours.

Patients were divided into the following groups based on cause of injuries as a result of Car Accidents - 104 patients (21.9%); injuries as a result of Car Collisions - 103 patients (21.7%); injuries received as a result of the Fall-193 patients (40.7%); household injuries - 17 patients (3.6%); injuries resulting from falling off a bicycle - 11 patients (2.3%); injuries resulting from beating - 15 patients (3.2%); patients with other injuries

Table 1 Characteristics of data for all patients and in different groups

Group	N	Age (m±δ)	N/F (%/%)	Died (%)	Survived (%)
All patients	474	9,0±5	303/171 (63.9/36.1)	40 (8.4)	434 (91.6)
Group based on cause of injuries					
Car Accidents	104	11.4±5	60/44 (57.7/42.3)	14 (13.5)	90 (86.5)
Car Collisions	103	11.4±3.8	58/45 (56.3/43.7)	7 (6.8)	96 (93.2)
Fall	193	7±5	136/57 (70.5/29.5)	14 (7.3)	179 (92.7)
Other	74	7.9±4.7	49/25 (66.2/33.8)	5 (6.8)	69 (93.2)
Age groups					
Infants and children of early childhood	69	1.4±0.6	47/22 (68.1/31.9)	8 (11.6)	61 (88.4)
Children of Preschool age	80	4.1±1.0	45/35 (56.3/43.8)	8 (10.0)	72 (90,0)
Children of Junior school age	143	8.6±1.8	93/50 (65.0/35.0)	10 (7,0)	133 (93,0)
Children of the Senior school age	182	14.3±1.9	118/64 (64.8/35.2)	14 (7,7)	168 (92,3)

N -sample's size of group, m - average age in the group, δ - standard deviation of mean, M – man-child, F – female-child, Died – the number of deaths in the group of patients, Survived - number of surviving patients in the group.

- 31 (6.5%). The last 4 groups in the processing of data were combined into the group of the Other 74 patients (15.6%) (Table 1).

Due to the fact that the physiological parameters of the patients depend on the age (A.V. Mazurin, 1999) [14], grouping was performed: infants (up to a year) and children of Early childhood - 69 patients (14.6%); children of preschool age - 80 patients (16.9%); children of junior school age - 143 (30.2%); children of the senior school age - 182 (38.3%). Infants are combined with children of early childhood due to the scarcity of the first group (Table 1).

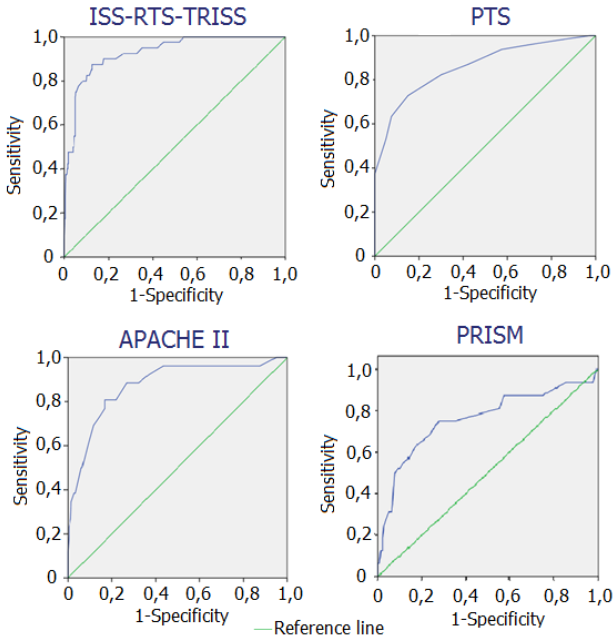


Figure 1. ROC-curves for prognostic scores.

Table 2 Results of AUROC and cut-off of prognostic scores.

Score	N	AUROC	Cut-off	Se	Sp
ISS-RTS-TRISS	474	0.928±0.020	0.149	0.900	0.823
PTS	474	0.859±0.025	4 score	0.850	0.728
APACHE II	399	0.875±0.038	0.168	0.808	0.834
PRISM	166	0.760±0.075	0.172	0.750	0.727

N -sample's size, AUROC- Area under ROC curve, cut-off- cut-off point, Se-sensitivity, Sp–specificity.

The assessment of severity of patient's condition and injury severity was done with prognostic scores PTS, ISS-RTS-TRISS, PRISM and APACHE II; and the death risk was determined using ISS-RTS-TRISS, PRISM and APACHE II scores; these data were collected from anamnesis, initial examination, parameters of physiological and neurological status on admission and within the first 24 hours. Scores were calculated using software tools published on the Sociéte Françaiseet de Réanimation community website [20]. The validity and prognostic accuracy of prognostic scores were assessed by determining its discrimination and calibration ability. The discrimination ability has shown a probability score to divide patients into two groups: with favorable outcomes and unfavorable (likelihood and lethal) outcomes. ROC-curve was used for evaluating the discrimination ability of the studied prognostic score [15, 18]. For searching the optimal cut-off point, we used the Youden Index (YI) [19]. The Hosmer-Lemeshow test (C-criterion) was used for defining the calibration ability of the studied prognostic score [16].A comparison of the discriminatory ability of scores was carried out by comparing the areas under the ROC curves with the z-criterion (DeLong method) [17]. The data base statistical assessment and statistical description were done in Excel and SPSS programs.

3. Results

The investigated prognostic score ISS-RTS-TRISS showed outstanding discrimination ability (0.928) the investigated predictive scores PTS and APACHE II showed excellent discriminative ability (0.859 and 0.875) [18]. The investigated PRISM score showed acceptable discrimination ability (0.760) (Figure 1).

For searching the optimal cut-off point, we used the Youden Index (YI) and we observed a condition in which the total value of sensitivity and specificity was the maximum [19]. The best cut-off-point for ISS-RTS-TRISS obtained at the level of 0.149, for the PTS - 4 score, for the APACHE II score 0.168 and for the PRISM score - 0.172 (Table 2).

Table 3 The results of a pairwise comparison of the discrimination ability of the scores.

Score pairwise comparison	z	p
ISS-RTS-TRISS and PTS	**1,666**	**0,048**
ISS-RTS-TRISS and APACHE II	**1,851**	**0,032**
ISS-RTS-TRISS and PRISM	**2,268**	**0,012**
PTS and APACHE II	-0,278	0,609
PTS and PRISM	1,210	0,113
APACHE II and PRISM	1,487	0,068

z – value of z criterion test (DeLong method), p- p – value of z criterion test, at p <0.05 the null hypothesis, that the areas under the curves are equal, is rejected.

Table 4 Results of assessment of calibration ability for prognostic scores.

Scale	N	χ2	p
ISS-RTS-TRIS	474	13.1	0.107
PTS	474	19.52	<0.001
APACHE II	399	88.5	0.012
PRISM	166	8.37	0.398

N- number of patient in the study, χ2- χ2- value , p-p-value (Hosmer–Lemeshow test).

The results of a pairwise comparison of the discrimination ability of the scores by comparing the areas under the ROC curves have shown the presence of statistically significant differences between the ROC-curve areas when comparing the ISS-RTS-TRISS and PTS scores, ISS-RTS-TRISS and APACHE II, ISS-RTS-TRISS and PRISM (the area under the curve for the ISS-RTS-TRISS score is statistically significantly larger than the areas under the PTS, APACHE II and PRISM scores, p <0.05). In turn, in the other pairs of PTS and APACHE II, PTS and PRISM and APACHE II and PRISM, no statistically significant differences were found (Table 3).

C-criterion of the Hosmer-Lemeshow test showed satisfactory calibration ability for prognostic scores: for ISS-RTS-TRISS and PRISM (p equally 0.107 and 0.398, respectively). The calibration ability for PTS and APACHE II were unsatisfactory (p less 0.05) (Table 4).

The analysis of the prognostic accuracy and validity of the scores in the groups for the cause of trauma has shown that among the scores under study, the ISS-RTS-TRISS score has outstanding discrimination ability with satisfactory calibrating ability in 3 out of 4 groups due to trauma. The three remaining scores were able to show good results only in one group of injuries - The Other (the AUROC >0.900 and p-value of calibration ability>0.05) (Table 5).

The analysis of the prognostic accuracy and validity of scores in different age groups showed a similar trend: the ISS-RTS-TRISS score has outstanding discriminatory ability with satisfactory calibrating ability in 2 out of 4 groups: the Infants and Children of Early childhood and the Junior school age groups. At the same time, the remaining scores were able to show high results also in one age group - Junior school age (the AUROC >0.900 and p-value of calibration ability>0.05).

4. Discussion

Modern medicine in pediatric traumatology focuses on increasing the effectiveness of decisions taken during treatment, based on assessing the severity of the condition and the

Table 5 The groups in which the results of validity and prognostic accuracy of prognostic scores were satisfactory.

Score	Groups due to injuries (AUROC; p)	Age's groups (AUROC; p)
ISS-RTS-TRIS	TheCar Collision (0,962; 0,256) The Fall (0,917; 0,899) The Other (0,962; 0,980)	Infants and Children of Early childhood (0,970;0,657) Children of Junior school age(0,982; 0,561)
PTS	The Other (0,920; <0,05)	Children of Junior school age (0,989; <0,05)
APACHE II	The Other (0,908; 0,850)	Children of Junior school age (0,9941; 0,281)
PRISM	The Other (0,961; 0,703)	Children of Junior school age (0,929; 0,753)

AUROC- the areas under ROC curve, p-p-value (Hosmer–Lemeshow test).

severity of injury in patients, predicting the likelihood of death in patients. One of the methods for solving this problem is the use of numerous specialized prognostic scores for various nosologies and different age groups of patients. A large number of prognostic scores and their application in practice led to the need to study their effectiveness in the conditions of the national departments of pediatric intensive care and intensive care. An important point in the use of scores is the quality and validity of the predictive model, which determines the validity and suitability of the application of the technique in specific conditions, and consequently - the correspondence of the possibility of prognostic scores to the tasks assigned. In our research, we included four prognostic scores: ISS-RTS-TRISS, PTS, APACHE II and PRISM to assessment of the severity state of children with injuries in conditions of specialized trauma hospital. For study validity and prognostic accuracy of prognostic scores were assessed by determining its discrimination and calibration ability.

Thus, the result of the validity study on the predictive accuracy and validity of the four prognostic scores in children with trauma indicated to use only these two scores: ISS-RTS-TRISS and PRISM. The ISS-RTS-TRISS score showed significantly outstanding predictive accuracy (AUROC=0.928).

Two other prognostic scores APACHE II and PTS are not applicable for predicting the likelihood of death in children with trauma, due to unsatisfactory calibrating ability although they show sufficient discriminative capacity.

The results of a pairwise comparison of the discrimination ability of the scores was shown the significant differences between the ROC-curve areas of ISS-RTS-TRISS and other scores ($p < 0.05$).

The analysis of the prognostic accuracy and validity of the scores in the groups for the cause of trauma and in the groups of different age trauma has shown that among the scores under study ISS-RTS-TRISS score may be useful in practice (has excellent discrimination ability and significant calibration ability in most groups).

Thus, as a result of our study, it was shown that only one of the four prognostic scores, ISS-RTS-TRISS, can be successfully used in the everyday practice in the department of anesthesiology and resuscitation in the specialized hospital of children's traumatology to assess the severity of the condition, with the possibility of predicting the likelihood of a lethal outcome.

5. References

[1] M. Letts, D. Davidson, P. Lapner ,Multiple trauma in children: predicting outcome and long-term results, Can. J. Surg. – 2002. - Apr;45(2). –P.126. - PMID: 11939656.

[2] P. Bhadoria, A.G. Bhagwat, Severity Scoring Systems in Paediatric Intensive Care Units, Indian Journal of Anaesthesia.- 2008.-52:Suppl (5).- P.663.

[3] E. Cantais, O. Paut, R. Giorgi, L.Viard, J. Camboulive, Evaluating the prognosis of multiple, severely traumatized children in the intensive care unit, Intensive Care Med. -2001.- Sep;27(9).- P.151.-

[4] H. R.Champion, Trauma scoring/ H. R. Champion// Scandinavian Journal of Surgery.- 2002. - 91(1). – P.12.

[5] C.K. Senkowski, M.G. McKenney, Trauma scoring systems: a review, J American College of Surgeons -1999, Nov; 189(5): 491-503, Qureshi,

[6] S.P. Baker et al. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care, Journal Trauma14 (1974), 187-96.

[7] C.R. Boyd et al. Evaluating Trauma Care: The TRISS Method, Journal Trauma 27 (1987), 370-8.

[8] KnausW.A., Drapper E.A., Wagner D.P., et al. APACHE II: a severity of disease classification system, JournalCritical Care Medicine 13 № 10 (1985), P. 818–829.

[9] M.M. Pollack, E. Ruttimann, P.R. Getson, Pediatric risk of mortality (PRISM) score. Journal Critical Care Medicine; 11 (1988), 1110-1116
[10] J.J. Tepas, D.L. Molitt, J.L. Taibert and M. Bryant, The Pediatric Trauma score as a predictor of injury severity in the injured child, Journal Pediatric Surgery 22 (1987), 14-18
[11] E. J. MacKenzie, J. A. Morris, G.S. Smith , M. Fahey/ J. P. Marcin, M. M. Pollack Acute hospital costs of trauma in the United States: Implications for regionalized systems of care, J. Trauma. - 1990. - №30(9). P.1096.
[12] D. A. Martins, L. R. Santos, M. A. Figueiredo, M. D. Guerra, R. S. Magalhães, R. M. Francisco, C. A. Rebelo, C. A. Abreu, A. R. Laiginhas, M. O. Dias, J. G. Chaves; A. Costa-Pereira, R. Oliveira, Assessment and Optimization of prognostic scores in Portuguese PICUs, Students' Research Papers IntroMed. - 2009/2010. - p 73.
[13] I.V. Vasilyeva, S.V. Meseherakov, A.B. Arseniev, Z.B. Semenova, The prognostic scale CRASH in the treatment of children with severe traumatic brain injury, Stud Health Technol Inform. -2015. - P.213.
[14] A.V. Mazurin, I.M. Vorontsov, Propedeutics of childhood diseases: a textbook for students of medical. universities // St. Petersburg: Foliant, 2001. – P.926.
[15] M. Greiner, Two-graph receiver operating characteristic (TG-ROC): update version supports optimization of cut-off values that minimize overall misclassification costs, Journal Immunology Methods 191 (1996), 93-94
[16] S. Lemeshow, D.W. Hosmer, A review of goodness of fit statistics for use in the development of logistic regression models, Journal Epidemiology 115 (1982), 92- 106
[17] E. R. De Long, D. M. De Long, D. L. Clarke-Pearson ,Comparing the areas under two or more correlated receiver operation characteristic curves a nonparametric approach, Biometrics. 1988. - Sep. 44(3). - P.837.
[18] D. W. Hosmer & S. Lemeshow - Applied Logistic Regression - 2nd Edition 2000, New York: Wiley.
[19] Youden, W.J. Index for rating diagnostic tests / W.J. Youden // Cancer.– 1950. - Jan;3(1). –P.32.
[20] http://test-app.sfar.org/welcome-the-sfar-website/scoring-systems-for-icu-and-surgical-patients/ (Access during 2008-2016 years, last access 07.03.2018).

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-270

Feasibility and Design of an Electronic Surgical Safety Checklist in a Teaching Hospital: A User-Based Approach

Karin KIEFEL[a], Klaus DONSA[a,1], Peter TIEFENBACHER[b], Robert MISCHAK[c], Gernot BRUNNER[d], Gerald SENDLHOFER[b,d,*], Thomas PIEBER[a,e,*]
[a]*JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria*
[b]*Executive Department for Quality and Risk Management, University Hospital of Graz, Austria*
[c]*FH JOANNEUM, Graz, Austria*
[d]*Research Unit for Safety in Health, c/o Division of Plastic, Aesthetic and Reconstructive Surgery, Department of Surgery, Medical University of Graz, Austria*
[e]*Medical University of Graz, Austria*

Abstract. *Background:* The Surgical Safety Checklist (SSC) is routinely used in operating rooms (OR) but its acceptance is low. One promising way to improve acceptance of the SSC and thus quality of patient care is digitalization. *Objective:* To investigate how a digitalization of the SSC could be implemented in a teaching hospital. Based on the identified user requirements we designed a first user interface (UI). *Method:* We performed a literature review, identified user perceptions and requirements during 12 interviews including a standardized questionnaire in surgical departments at the University Hospital Graz (Austria). Subsequently a first prototype of a UI was designed. *Results:* Seven different approaches for digital SSC were identified in literature. Our interviews showed that 90% of the participants had a positive attitude towards a digitalization of SSC. The most favoured version of a digitalized SSC was a tablet-based client-server system with integration in the EHR and projection on an OR monitor. *Conclusion:* Digitalization of the SSC is requested by medical and nursing personnel. Based on the identified user requirements we designed a process oriented UI of a digital SSC.

Keywords. eHealth, computer aided surgery, patient safety, surgical procedures, checklist

1. Introduction

The rate of surgery associated complications lies between 3% and 16% in developed countries, with a worldwide mortality of 1 million deaths per year [1]. Surgery associated complications are mostly wrong patient, wrong side, wrong procedure or even retained surgical items. Studies have also shown that poor teamwork and insufficient communication during surgery cause around 43% of all surgical failures. In addition, the risk of complications in operating rooms (OR) is increased by inadequate pre-surgical preparation like antibiotic prophylaxes. Nearly half of all complications are avoidable [1-3].

[1] Corresponding Author: Klaus Donsa, JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, E-Mail: healthca@joanneum.at
* both authors equally contributed to this study

In order to prevent surgical complications, the WHO published the "Surgical Safety Checklist" in 2007 [4] which should act as a supporting tool for the surgical staff to maintain patient safety and promote a reduction of errors such as wrong patient, , wrong side, surgical infections, perioperative complications and deaths [5].

Numerous studies have already proven that the use of this Surgical Safety Checklist (SSC) has positive effects on safety during surgery [6-9]; decreased mortality by 47% and complication rate by 36%. Communication and team cooperation is also positively influenced by the SSC [10].

In 2011 a modified paper-based WHO SSC was implemented within the University Hospital of Graz (UHG). Since then it became a mandatory tool in all surgical departments, but overall acceptance is still not reached yet. The use of the SSC has changed positively over the past years with nearly 90% now using the SSC. Nevertheless, the satisfaction with the SSC is comparatively low (only 55% of the nurses are satisfied or very satisfied) and completion rates drop down over the time (from 81.7% to 57.2%) [11,12]. In addition to that, the opening of the new central surgery at the UHG in October 2017 has a great impact on coordination and safety of patients, which add additional value to the use of a SSC. One promising way to improve acceptance of the SSC and thus quality of patient care and its safety is digitalization [13,14].

This article deals with the question if and how a digitalization of the SSC could be implemented in a teaching hospital. Literature was reviewed and user perceptions and requirements were identified and were used to design a digitalized SSC.

2. Methods

We performed a systematic literature review and qualitative (interviews) and quantitative (questionnaires) analyses to identify user perceptions and requirements. Based on the results we designed a first prototype of an UI.

2.1. Systematic literature review

We conducted a PubMed search (October 2017) focusing on the phrases "*surgical*", *(*or *"surgery", "surgical procedures", "perioperative")* in conjunction with *"checklist" (*e.g. *"safety checklist", "surgical checklist")*, "*patient safety*" and "*electronic checklist*" (or "*digital system*", "*digital assistance*", "*computerized*"). We additionally applied the snowball principle and categorized the identified literature by relevance.

2.2. Qualitative analysis of interviews

For qualitative analysis, we conducted 12 interviews. One half of the questions related to the current use of the paper-based SSC (satisfaction, problems and weaknesses of the SSC) and the other half dealt with possible improvements of the SCC and implementation conceptions towards digitalization. The participants have been selected by the Quality and Risk Management of UHG and came from different professional groups and departments to cover a wide range of SSC user groups (Table 1). During the interviews answers were documented in writing and audio recording. Afterwards, we analyzed all interviews by summarizing and categorizing the information.

Table 1. Interview participants (MF=managerial function)

Professional Group	Department	Position
Surgeon	Otolaryngology	Expert with MF
	Traumatology / Orthopedics	Expert with MF
	Plastic Surgery	Expert with MF
	Transplant Surgery	Expert without MF
	Thoracic and hyperbaric surgery	Expert with MF
	Special anesthesiology	Expert with MF
Assistant of surgeon	Cardiac-, thorax-, vascular surgery, anesthesiology and intensive care medicine	Expert without MF
Nurse	Otolaryngology	Expert without MF
	General surgery	Expert without MF
Anesthesia nurse	Special anesthesiology	Expert with MF
Surgery manager	Executive department of surgery management	Expert without MF
	Executive department of surgery management	Expert with MF

2.3. Quantitative analysis of questionnaires

For quantitative analysis, a standardized questionnaire with 16 questions was filled out by each participant after the interview (participants listed in Table 1). Opinions towards the currently used SSC and a digitalized checklist implementation, system requirements and suggestions for a possible digitalized SSC were inquired. The full list of questions can be requested. The participants had to pick the matching scale boxes of an ordinal 5 point Likert Scale, a nominal form ("*yes*" / "*no*" boxes) as well as pre-formulated selection options.

2.4. Iterative design of the user interface

Based on results from the literature review and focusing on specific user requirements which were gained from the interviews and the questionnaire, a first prototype UI of a new digital SSC was iteratively developed. This intends to pass a cyclic process of 4 main phases (Inception, Elaboration, Construction and Transition) with continual quality and functionality improvements after each of the iterations. The UI was then further refined according to improvement suggestions from users in the University Hospital of Graz and project members of JOANNEUM RESEARCH.

3. Results

3.1. Literature review

Thirteen out of 189 articles remained after exclusion of not relevant results. Reasons for exclusion were no relation to digitalized checklists, no relation to use in the surgery, duplicates or language other than German and English.

We used the included articles to compare evidence about electronic/digitalized SSC. Three publications showed the same checklist or a modified version of the checklist. One was a literature review about electronic checklists [15] and in one article development of an individual digital compilation of electronic checklists was described and a possible implementation in hospitals was suggested [16]. Seven different types of digital SSC exist and were summarized in Table 2. Main findings concerning features and benefits were: six out of 9 identified approaches of digital SSC were using monitors in OR during

team time out (TTO) with additional access to EHR (4 with interactive functionality). Furthermore, some of them support functions like critical information messages (auto-populations), process stops when items were not completed and process oriented workflow visualization. The most frequently achieved benefit was increased user compliance (6 of 9) and increased patient safety (6 of 9).

3.2. Qualitative analysis

Qualitative analysis based on 12 interviews revealed that nearly all interviewed persons confirmed that the currently used SSC is essential for patient safety. Seven participants said that the checklist is finally accepted but it took a long time, the remaining five participants added that the importance, meaningfulness, efficiency and responsibility are often not seen by the medical staff. A main aspect for poor user acceptance was the redundancy of items to be checked. The interview partners suggested that the SSC has to be redesigned or reduced, that a clear definition of responsibilities has to be implemented and that trainings have to be organized to achieve a better understanding of checklist items.

Reported problems of the SSC were: not all team members were present at TTO, the "moment of rest" was not enforced (attention is low) and additional time effort is produced if processes are not familiarized. To avoid these problems, the participants emphasize following points: the SSC should be short, augment the culture of safety (elucidate benefits and consequences, conduct trainings), the design should be process oriented (procedural, seamless handover, information gathering and communication flow).

Table 2. Results of the literature review: features and benefits of electronic checklists.

SSC type	Features	Benefit
Electronic flight board with clinical decision support (CDS) [14]/[17]*	Real-time patient data (EHR), critical information (auto-populating), process stop when item incomplete, access and projection of data (EHR), procedural, client-server-system (CSS)	Increased compliance, increased patient safety and staff acceptance
Pre-recorded audio checklist [18]	Audio delivery of items	Increased compliance, consistency of questions and staff attention
Pre-recorded audio checklist (mobile device connected to EHR [13])	Interactive screen, procedural, Real-time patient data (EHR), CSS	Increased compliance, staff acceptance and improved workflow efficiency
Video-based checklist [19]	Access and projection of data (EHR), audio delivery of items, CSS	Increased compliance, increased patient safety, staff acceptance and improved workflow efficiency
Interactive screens (LCDs linked to EHR) [19]/[20] **	Real-time patient data (EHR), access and projection of data (EHR), process stop when item incomplete, procedural, progress is visualized, CSS	Increased compliance, increased patient safety, improved workflow efficiency, cost saving
Integrated in OR connected with interactive screen (and EHR) [21]/[22]*	Real-time patient data (EHR) and critical information (auto-populating)	Increased compliance, increased patient safety, improved communication, improved workflow efficiency, cost saving
Integrated in OR connected with interactive screen [23,24]	Access and projection of data (EHR), procedural, progress is visualized	Increased patient safety, improved staff acceptance and communication, time saving

** same checklist different paper/method; ** same checklist but modified*

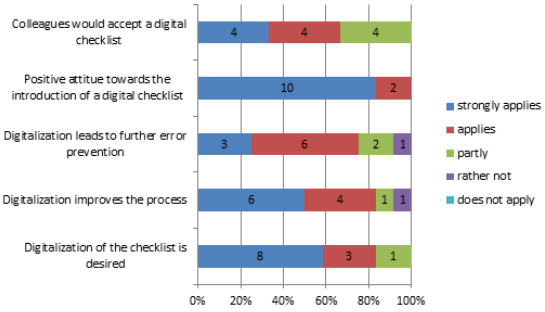


Figure 1. Results from the standardized questionnaire (excerpt).

Further requirements to an electronic SSC with strong agreement were: improved hard-stop culture (prohibit further steps), improved "moment of rest" (audio signal during TTO), prevent incompleteness (required fields), improved quality of data (no loss of data, better archiving and traceability), time saving (more efficient, has fewer redundancies, complete documentation), increase safety (warnings, no loss of device, less manipulation) and improves the workflow (less paperwork, better archiving, documentation, comprehensibility and evaluation).

The participants had a positive attitude towards a mobile solution in form of a tablet connected with a computer or monitor. Emphasis has to be placed on a user-friendly design. Eight out of 12 interview partners favored the idea to have the checklist information projected on a monitor in the OR. A connection to EHR is required from the majority (9 out of 12). A new digital SSC has to focus on seamless workflow integration. Following functions supported by a new digital SSC should be considered in a future development according to the interview partners: automatically retrieve patient data by scanning a QR-code on the patient wristband; display relevant notes, display patient images, automatically include diagnosis and patient data, scan barcodes of consumed materials during the operation (for documentation), handle postoperative arrangements and notes (free text), implement warnings and reminders (for allergies, antibiotic prophylaxis, incompleteness of items).

3.3. Quantitative analysis

Quantitative analysis based on the evaluation of questionnaires revealed that 7 out of 12 of the participants were satisfied with the currently used SSC, but stated that a further development on the paper-based checklist is necessary. Reasons were that the use is not mandatory, the use is not uniform, a signature for accountability is missing, there are ambiguous formulations of checklist items, some items are not suitable, the process is not displayed adequately and the current design should be improved.

Only 33% agreed with the statement that the responsibilities were reasonably regulated with the SCC. Main reasons for the low favorable result were irresponsible completion of the SSC and the hierarchic culture in OR. The opinion towards that a consistent use of the SSC can prevent errors was agreed by all participants (67% voted "*strongly applies*", 33% voted "*applies*") but some commented "only if used properly". Only half of the participants reported that the communication and information access regarding the SSC worked seamlessly.

Nearly all participants agreed with the statement that complications can be avoided due to the use of a SSC. 'Operation location marked' and 'patient identification' were

Figure 2. Prototype of the UI (tablet) for the new digital SSC.

the two most important items on the current SSC according to interviewed participants. Nearly all of the participants had a positive attitude towards a digital SSC implementation and more than half of participants evaluated that a digitalization of the checklist is strongly desired with an overall agreement that a digitalized process could improve the workflow, see Figure 1. A new digital SSC in the form of a tablet and/or a computer in the OR connected to EHR was favored by the majority (9 of 12).

3.4. User interface design

Based on the requirements from the literature review and from qualitative and quantitative analyses, a prototype of a UI was designed (Figure 2). We focused on an easy and user-friendly layout and supported a process oriented use of the SSC. Patient information will be displayed during the whole checklist process.

We chose a rugged tablet device for the usage in OR (Samsung Galaxy Tab Active). The UI prototypically provides features like: scan patients wristband (assures correct patient identification), scan barcodes of materials (managing consumption of materials during operation) and login of different users.

Users will be guided through the entire process by procedural visualization (e.g. SIGN IN is divided in nursing and anesthesia personnel). Auto-populating warnings remind the users of allergies and barriers should block the user from proceeding if the process is incomplete (TTO is only available if SIGN IN is completed). The connection to the EHR and to computers in the OR is envisaged. Interview participants preferred a tablet-based solution instead of a standard computer because they have to check SSC items immediately at point-of-care. There will be no additional infrastructure required to implement the solution within hospital settings because the equipment is already available (WiFi, wall-mounted displays, computers, speakers and barcode scanners).

4. Discussion

Our findings revealed that an electronic SSC is desired by the majority of interview partners. Based on these promising results, we designed a prototype UI for a digitalized

SSC. The comment: "a fool with a tool is still a fool" emphasize the critical attitude of surgical team members towards an electronic application. Without an added value, there will not be any changes in active participation and checking the list properly. Therefore, many interviewed participants requested additional features such as patient identification over wristband (scanning QR code), including material documentation (scanning barcode) and free-text fields to add documentation. Functions like auto-populating messages, process oriented workflow, barriers to prohibit continuing in case of unfinished tasks and monitor projection during TTO were requested.

The integration into the hospital information system (HIS) is regarded as the best choice in order to raise compliance concerning SSC items [13]. However, the approach of a tablet-based digitalization together with a HIS integration is assumed to best support OR-team members in a ubiquitous way. With a CDS, e.g. automatically retrieving patient demographics, updated laboratory values, allergies, medications and audio signaling helps to increase attention in each of the complex steps (SIGN IN: entering the OR – SSC on tablet could be used at point-of-care, TTO: skin incision – SSC can be projected on monitor via HIS; SIGN OUT: before skin suture).

Main barriers for implementation of SSC are challenges regarding efficiency (double checks), lack of knowledge about the correct use and stringent hierarchical structures [2,11,25]. Design issues, lack of process integration and inefficient timing of checklist use is also problematic [26]. Today there are still problems with the overall acceptance and compliance with SSC [2,11,12]. Encouraging healthcare professionals with new tools is difficult as it is always associated with the change of habits and engagement [11,26]. New tools "requires the willingness of healthcare professionals" [11] and it is "an ongoing challenge towards the goal of gaining acceptance amongst healthcare professionals and raising compliance" [12]. However, recent developments that handle with an electronic version of a SSC reported an increase in compliance as well as improved patient safety [7,13,27-29]. We found no studies of SSC used on tablets-PCs in our literature search.

Problems regarding electronic systems mentioned in the interviews were the risk of a breakdown (due to damage), transmission-, operation- and documentation errors and slow systems. Additional challenges mentioned were: multidisciplinary documentation, coordination of trainings and acceptance issues.

In conclusion, we found that digitalization of the SSC is feasible and desired by the OR staff. Beside the general consensus regarding the usefulness of checklists in OR the majority of interviewed participants agreed that transferring the paper-based checklist into an electronic version could improve the whole surgical process. One design favored by the majority of interview participants was the implementation of an electronic SSC in a tablet-based CSS integrated in the EHR with projection to an OR monitor. We designed a first prototype UI which will be further expanded as functional demonstrator of the SSC application and then validated in future usability tests.

References

[1] T. G. Weiser et al., "An estimation of the global volume of surgery: a modelling strategy based on available data," Lancet, vol. 372, no. 9633, pp. 139–144, 2008.

[2] D. S. McDowell and S. A. McComb, "Safety Checklist Briefings: A Systematic Review of the Literature," AORN J., vol. 99, no. 1, p. 125–137.e13, 2014.

[3] WHO, "Safe Surgery Saves Lives," 2008.

[4] WHO Patient Safety. and World Health Organization., "Implementation manual WHO surgical safety checklist 2009 : safe surgery saves lives," p. 16 p., 2009.
[5] A. Haynes, T. Weiser, W. Berry, and Et Al, "A surgical safety checklist to reduce morbidity and mortality in a global population.," N. Engl. J. Med., vol. 360, no. 5, pp. 491–9, 2009.
[6] T. G. Weiser, A. B. Haynes, G. Dziekan, W. R. Berry, S. R. Lipsitz, and A. A. Gawande, "Effect of A 19-item surgical safety checklist during urgent operations in a global patient population," Ann. Surg., vol. 251, no. 5, pp. 976–980, 2010.
[7] W. A. Van Klei et al., "Effects of the introduction of the WHO 'surgical safety checklist' on in-hospital mortality: A cohort study," Ann. Surg., vol. 255, no. 1, pp. 44–49, 2012.
[8] P. Helmiö, K. Blomgren, A. Takala, S. L. Pauniaho, R. S. K. Takala, and T. S. Ikonen, Towards better patient safety: WHO Surgical Safety Checklist in otorhinolaryngology, vol. 36, no. 3. 2011.
[9] D. M. Conley, S. J. Singer, L. Edmondson, W. R. Berry, and A. A. Gawande, "Effective surgical safety checklist implementation," J. Am. Coll. Surg., vol. 212, no. 5, pp. 873–879, 2011.
[10] A. Fudickar et al., "The effect of the WHO surgical safety checklist on complication rate and communication," Dtsch. Arztebl. Int., vol. 109, no. 42, pp. 695–701, 2012.
[11] G. Sendlhofer et al., "Implementation of a surgical safety checklist: Interventions to optimize the process and hints to increase compliance," PLoS One, vol. 10, no. 2, pp. 1–14, 2015.
[12] G. Sendlhofer et al., "The gap between individual perception and compliance: A qualitative follow-up study of the surgical safety checklist application," PLoS One, vol. 11, no. 2, pp. 1–11, 2016.
[13] R. N. Uppot et al., "A Verbal Electronic Checklist for Timeouts Linked to the Electronic Health Record," J. Am. Coll. Radiol., vol. 14, no. 10, pp. 1322–1325, 2017.
[14] M. E. Gitelis et al., "Increasing compliance with the World Health Organization Surgical Safety Checklist—A regional health system's experience," Am. J. Surg., vol. 214, no. 1, pp. 7–13, 2017.
[15] H. S. Kramer and F. A. Drews, "Checking the lists: A systematic review of electronic checklist use in health care," J. Biomed. Inform., vol. 71, pp. S6–S12, 2017.
[16] N. Estima, "A New Electronic Approach for the Surgical Safety Checklist," 2014.
[17] T. Shear et al., "Procedural Timeout Compliance Is Improved With Real-Time Clinical Decision Support," J. Patient Saf., vol. 16, no. 0, p. 16, 2015.
[18] S. Reed, R. Ganyani, R. King, and M. Pandit, "Does a novel method of delivering the safe surgical checklist improve compliance? A closed loop audit," Int. J. Surg., vol. 32, pp. 99–108, 2016.
[19] R. Mainthia et al., "Novel use of electronic whiteboard in the operating room increases surgical team compliance with pre-incision safety practices," Surgery, vol. 151, no. 5, pp. 660–666, 2012.
[20] J. M. Ehrenfeld, N. Patel, W. S. Sandberg, and D. Ph, "Electronically Mediated Time-out Initiative to Reduce," 2018.
[21] Dan C. Krupka; Peter J. Logue, "Model-Business-Case-for-LiveData-OR-D-ATO," 2009.
[22] J. Robbins, "Hospital checklists transforming evidence-based care and patient safety protocols into routine practice," Crit. Care Nurs. Q., vol. 34, no. 2, pp. 142–149, 2011.
[23] S. N. Buzink, L. Van Lier, I. H. J. T. De Hingh, and J. J. Jakimowicz, "Risk-sensitive events during laparoscopic cholecystectomy: The influence of the integrated operating room and a preoperative checklist tool," Surg. Endosc. Other Interv. Tech., vol. 24, no. 8, pp. 1990–1995, 2010.
[24] M. L. McCarroll et al., "Development and implementation results of an interactive computerized surgical checklist for robotic-assisted gynecologic surgery," J. Robot. Surg., vol. 9, no. 1, pp. 11–18, 2015.
[25] P. S. Ragusa, A. Bitterman, B. Auerbach, and W. A. Healy, "Effectiveness of Surgical Safety Checklists in Improving Patient Safety," Orthopedics, vol. 39, no. 2, pp. e307–e310, 2016.
[26] S. J. Russ et al., "A qualitative evaluation of the barriers and facilitators toward implementation of the who surgical safety checklist across hospitals in England," Ann. Surg., vol. 261, no. 1, pp. 81–91, 2015.
[27] J. L. Dixon, D. Mukhopadhyay, J. Hunt, D. Jupiter, W. R. Smythe, and H. T. Papaconstantinou, "Enhancing surgical safety using digital multimedia technology," Am. J. Surg., vol. 211, no. 6, pp. 1095–1098, 2016.
[28] E. de Vries et al., "Effect of a comprehensive surgical safety system on patient outcomes.," N Engl J Med., vol. 363, no. 20, pp. 1928–1937, 2010.
[29] E. K. P. Mayer et al., "Surgical Checklist Implementation Project: The impact of variable WHO Checklist compliance on risk- adjusted clinical outcomes following national implementation. A longitudinal study," vol. 53, no. 9, pp. 1689–1699, 2013.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-278

Using MEESTAR to Identify Ethical and Social Issues Implementing a Digital Patient-Centered Care Platform

Johannes WUTZKOWSKY[a,1], BRITTA BÖCKMANN[a,b]
[a] *Department of Computer Science, University of Applied Sciences and Arts, Dortmund, NRW, Germany*
[b] *Institute of Medical Informatics, Biometry und Epidemiology (IMIBE), University Hospital Essen, Germany*

Abstract. The PIQ research project ("Pflege im Quartier") aims for optimizing communication and patient-centered care for elderly people by implementing a digital, patient-centered care platform, with particular attention is paid to the consideration of ethical, legal and social issues. In this work, an instrument for the ethical evaluation of social-technical arrangement was used to map features of the platform with ethical concerns. The results include precise ethical questions based on scenarios identified and possible solutions to address them, regarding the ethical and privacy issues. These insights will be continuously integrated in the future system design and implementation as well as research.

Keywords. elderly; independent living; home nursing; ethics

1. Introduction

The demographic change in the Federal Republic of Germany which is accompanied by social challenges (e.g. the predicted shortage of professional nurses and the increasing need for care) which are to be encountered by innovative technology in the area of health and long-term care. For this purpose, research projects are funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF), which contribute to the development and research of innovative approaches to support people in need of care and people involved in the medical and nurse health care. One goal of these new technological innovations and systems is to enable elderly people in need of care and assistance, to live a self-determined and independent life in their homes and in a familiar environment [1]. This leads to reduce costs, relieve caregivers (professional, informal) and contribute towards efficient networking.

In particular, technology that is designed and developed for elderly, who are or will be in need of care and are therefore more vulnerable, must be ethically evaluated to guarantee a harmless use, as well as it can contribute to the user-acceptance, and reduce the market entry barriers [2]. It is important that the ethical evaluation accompanies the development process from the beginning to ensure potential advantages and reduction of risks.

[1] Corresponding Author: Johannes Wutzkowsky, University of Applied Sciences and Arts, Department of Computer Science, Emil-Figge-Straße 42, 44227 Dortmund, E-Mail: johannes.wutzkowsky@fh-dortmund.de

In this contribution, we present recent work-in-progress from the BMBF-funded project PIQ (2016-2019) where scientists are accompanied by ethical experts that contribute to identifying and assessing ethical conflicts and their relevance. The purpose is to establish a systematic approach using a model for the ethical evaluation of socio-technical arrangement (MEESTAR) [3]. We identified a collection of ethical/ moral conflicts and principles that take a significant role in the development of a technical assistance system. Next to the identification of conflicts, we develop potential solutions.

1.1. The Project PIQ (Pflege im Quartier): general information

The main approach of the project deals with the information, communication and networking around nursing for people in need of care, their relatives, certified nurses and informal caregivers. The development and use of technical assistance systems in the everyday life of elderly people who may be in need of care and because of their cognitive abilities or other restrictions are no longer able to reflect the scope of their decisions or do not understand the full functionality of the system, comprehensible explanations and an intuitive use of the system are essential to avoid excessive demands.

The aim of the project is to optimize and build up real and digital structures that exist in four heterogenic quarters in the city of Gelsenkirchen in North Rhine-Westphalia, Germany. The optimization includes the improvement of the medical and nursing health care by supporting the home care of people in need to live an independent and self-determined life in their own preferred environment [1]. Furthermore the professionals are supported by expanding the integrative supply through a demand-meeting information exchange that contains structured data and is based on established medical standards, such as HL7 CDA based electronic nursing record [4]. This supports the comprehensive networking between all those people and institutions involved especially as part of the care transition and hospital discharge [5].

1.2. The Project PIQ: platform and mobile application

The digital components that are developed within the project include a mobile application and a patient-centered platform for people in need of care, their relatives and people that are involved in the medical and healthcare process.

One aspect of the platform includes the possibility to locate people in certain situations through the tracked location by their phone if the mobile application is installed. Possible situations are e.g. if they are far away from their supposed location over a longer period or got lost. On one hand, a set of rules can be defined by the caretakers and confidants that are authorized to keep track of the current whereabouts of their relatives or people in need of care and on the other hand an emergency call/ alarm can be triggered manually by a “Help through the touch of a button”- function.

The mobile application offers features to import a printed medication plan (bundeseinheitlicher Medikationsplan) and to organize the medication within. The information given by the medication plan is automatically transformed into a structured format, so that it can be stored on the mobile device. The structure of the medication plan relies on the technical specification that is part of the pharmacotherapy safety to provide a consistent and structured format for the electronic management, storage and cross-system exchange of the currently prescribed medication. The goal of integrating and exchanging the medication plan is to prevent adverse effects by recognizing and reviewing drug interactions and contraindications and to contribute to the

pharmacotherapy safety [6,7]. The medication plan contains information about the prescribed medication that a specific person is currently taking. This functionality enables the people in need of care and their relatives to organize and their medication management and support their safety.

Another feature is the integrated marketplace in the system. It allows people to get in touch with volunteers and neighbours in their quarter to increase participation in social life. It is possible to offer and search for abilities and help as well as for support of activities in their everyday life that can then be arranged. The use of this feature is optional, which guarantees the right not to participate. There will be new digital advisory services that include new formats to involve people with voluntary commitment as well as training content related to caretaking for the relatives based on the care-level and the current situation of the people in need of care.

The paper is organized as follows: First, in section 2, we will describe the selection of the evaluation model and how it can be integrated into the software development and workflow process. Section 3 describes the intermediate results consisting of selected scenarios and developed solutions. In section 4 a summary of the work will be presented and demonstrates which future work will have to be done.

2. Methods

When deciding which framework for the ethical evaluation should be used, it is important to consider the appropriateness for the particular context, the scope and purpose of the ethical analysis as well as the way in which the instrument addresses issues within the considered domain. The provided theoretical, ethical principles and conflicts have possible limitations by their generalization and transferability or applicability.

The system that is developed includes features that are intended to support people in need of care in their everyday life through localization and orientation functions, medication management, care-related information and communication and networking options to participate in social life. An instrument for the ethical evaluation should be practically (normatively) usable and integrable into the process of designing and developing an information and communications system.

In contrast to other guidelines and ethical evaluation instruments such as “a model for the ethical evaluation of socio-technical arrangement” (MEESTAR), the “Model of Assessment of Telemedicine” (MAST) or the VDI Guideline 3780. MAST was developed for telemedicine applications, while the VDI guideline is for the evaluation of technology in a more general sense that enables a normative assessment but provides no clear methodology how it could be performed [8]. In contrast, MEESTAR provides a clear methodology and enables a normative assessment. For these reasons, the MEESTAR model has been selected as a suitable tool to conduct the ethical reflection.

2.1. The MEESTAR-Model

MEESTAR offers an approach to identify and describe ethical issues and dilemmas that can be assigned to given ethical dimensions. In addition, it provides questions for each dimension that can be used as a basis for a discussion and evaluation. The analytical model is divided into three different axes that need to be considered (Figure 1). One axis shows the three different points of view from which the socio-technical system is examined by: the individual level, the organizational level and the social level. For each

of these levels an evaluation takes place which includes seven ethical dimensions: care, autonomy, safety, justice, privacy, participation and self-conception. The third axis contains four stages with a different ethical sensitivity that the functionality or the ethical topic is categorized to. These stages comprise the completely harmless use (stage I), the ethical sensitivity that can be compensated in practice (stage II), the ethical extreme sensitivity that requires either permanent monitoring or the introduction should be questioned (stage III) and the rejection of use (stage IV).

The use of the model for evaluating an information and communication system can contribute to highlight potential ethical conflicts that arise from the use of technology in nursing and medical health care that should be considered. A situation is assigned to a dimension and viewed from the specific levels (e.g. the individual level). After careful assessment of the existing conflicts an ethical stage is assigned. In addition, it provides ethical guidelines that should be considered to create the system. The model does not offer a definitive assessment of the system, it rather contributes to the considerations of ethical and moral conflicts.

2.2. Integrating MEESTAR into the software development and workflow process

The system is developed according to a development process model, which consists of six different steps (Figure 2). MEESTAR was integrated and applied after the system requirements were defined to identify ethical aspects and develop appropriate measures as soon as possible and to include them in the conception and implementation process of the system. MEESTAR provides a procedure for the ethical evaluation that is often used as a basis for discussions in workshops. We deviated from that process and we slightly modified the proceeding. For this reason, we applied MEESTAR in the development process as follows. The model was used to identify the ethical and moral conflicts and extended to reveal more specific questions that can be met by possible solutions.

Step 1: Together with involved people in the medical health care system or in the caring process (e.g. nursing service, nursing bases, public health care personal, people in need of care, relatives etc.) from the selected areas demands and requests are determined.

To achieve this, semi-structured interviews and the attendance of the mobile care process were performed. The requirement analysis is done by an interdisciplinary team.

Step 2: The gathered information and results can be used to define problems and scenarios to develop requirements for the system. In addition to the legal and technical

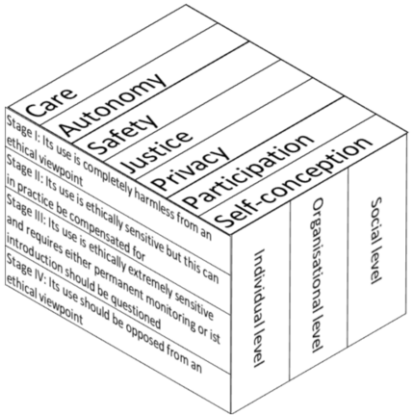


Figure 1: MEESTAR-Model Source: Own representation based on ([3], p.14)

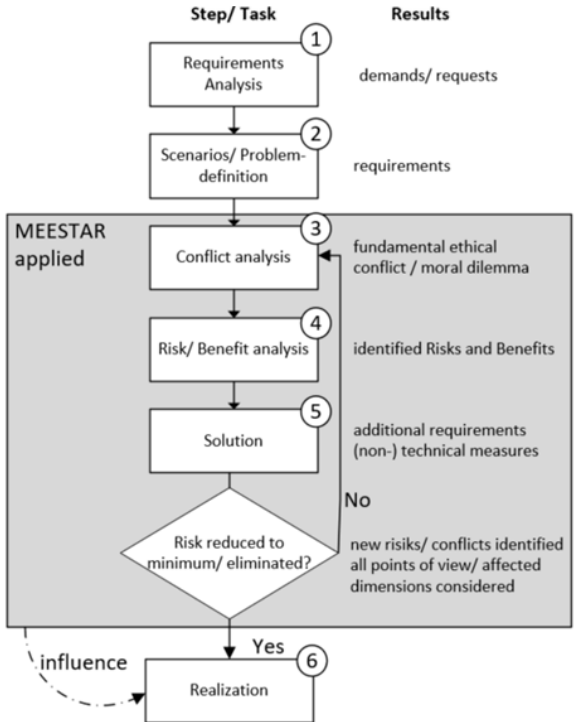


Figure 2: Integrating MEESTAR in the development

specifications and possibilities, it is important to identify ethical and moral conflicts/ dilemmas.

Step 3: To frame the ethical/ moral conflicts and dilemmas, some scenarios are analyzed by taking a particular perspective (individual, social, organizational), the affected dimensions, and related questions provided by MEESTAR into account.

Step 4: The affected dimensions and the identified moral/ ethical conflicts offer a possibility to focus on a specific aspect and to concretize risks and benefits. These can be compared to each other to make fundamental decision to define possible solutions.

Step 5: In this case, a critical/ objective position to the system must be taken and all possible points of view and different interests must be considered to ensure that the use of the system guarantees greatest possible safety and security. The developed solutions can be technical and non-technical as well as new requirements for the system.

Decision: Because of the different complexity of the identified problems the process can be repeated several times. This is done until it has been ensured that the risks have been reduced to a minimum or are eliminated, no further risks or conflicts have been identified and all affected dimensions, perspectives and interests have been considered.

Step 6: In the final step, the established solutions and measures are realized and continuously monitored and evaluated.

3. Results

Currently the project PIQ has reached mid-term. Hence, the system is in development and preliminary results are presented in this section, including the identified scenarios,

Table 1. Ethical conflicts and developed solutions (Assigned stage refers to Figure 1)

Dimension	Ethical question/ Ethical conflict	Developed solution	Assigned stage
Autonomy	How can a comprehensible provision of access rules and permissions be provided?	Simplified list of questions that binds control- and security settings to pre-configurated profiles.	I
	How to acquire the ability to consent?	Gain the user consent on basis of an informed consent discussion. If necessary relatives/ authorized representatives can be present.	I
	How to handle and grant access to nurse-related medical information?	In depth and simplified configuration of profiles for access-control and access-policies.	I
Safety	Could displaying the drug safety assessment lead to uncertainty?	Autonomy has to retain within the responsibility of the doctor/ pharmacist. The functionality is not implemented.	IV
	Could the marketplace as a possibility for networking and participation be abused to harm people in need?	The networking feature is supervised and controlled by an involved organization. Participation is optional.	III
	Could the localization function that is meant to support safety and mobility also restrict freedom and independence?	Functionality can be (de)activated. Transparency through logging. Reason for localization must be stated and is recorded.	II
Self-Conception	Can users develop a dependency on the functionality of the system?	Must be further evaluated.	
	Can the functionality cause possible behavioral adjustments?	Must be further evaluated.	

the development of precise solutions and countermeasures that will be included and realized. This was achieved using the software development process introduced in section 2.2. Additionally, an interdisciplinary workshop, which consisted of representatives from the fields of ethics, medical informatics and social science was performed to validate findings.

3.1. Identified scenarios

A collection of eight scenarios was identified and evaluated with the approach presented in section 2.2 (Table 1). Each of the scenarios and related ethical conflicts result in a different set of benefits and risks that needs to be encountered with appropriate measures. Solutions were created for each of these conflicts and each will be monitored during the ongoing research, e.g. through focus groups, expert discussions and field test.

3.2. Example Scenario: Localization

As an example, the process of evaluating the localization function is shown (Table 2). The system provides the feature to locate and determine the whereabouts of an affected person. Along with increased safety by localization in emergency situations, the function could be abused for controlling and surveillance. These consequences can be traced back to conflicts that concern the dimensions autonomy, self-conception, privacy, safety and care. The MEESTAR-Model provides different questions that help to identify the ethical conflict related to these dimensions. These questions were substantiated and derived to receive more concrete questions that can be encountered by possible solutions. This process leads to support of the user with comprehensive security and control features to remain in control of his personal information. Solutions were developed to prevent and minimize the improper use and to counter the negative risks of the offered functionality.

Table 2. Evaluating the localization function (Assigned stage refers to Figure 1)

Demand/ request	"When I'm on the road, I want to be sure that someone knows where I am."
Requirement	Current location is tracked by smartphone and transmitted to the platform where it can be accessed by authorized people only.
Dimension(s)	Autonomy, Self-conception, Privacy, Safety, Care
Ethical conflict	"[…] at which point does a well-intended caregiving attitude become a patronizing or negatively paternalistic approach […]?" ([3], p.15) "How do we resolve conflicts between safety and privacy and between safety and autonomy (freedom)?" ([3], p.16) "How can people be assisted in their autonomy on the basis of practices consistently throughout the individual's right to autonomy?" ([3], p.15).
Substantiated/ derived ethical questions	Who profits from the surveillance system? How does the safety/ surveillance affect the behaviour of the people involved? Does the localization serve the safety of the affected person or the certainty of the relative/ localizer?
Risks	Excessive use by the user; Inappropriate use (surveillance) by authorized people; limitation of personality development through the surveillance system
Benefits	increase mobility, security; support participation in social life; enhance independence
Solution	The functionality can be (de-) activated. Transparency is established by complete and accessible log entries. A reason for a performed localization must be given. Freedom to choose and use technology. Authorization can be configured for a specific time range and for specific persons.
Ethical stage	II ("its use is ethically sensitive but this can in practice be compensated for")

3.3. Further aspects

To increase the self-determination and autonomy of the user an informed consent discussion will take place. This procedure will be designed to achieve transparency and clarify that the system does not patronize, but rather promotes the digital competence and autonomy of the user before they give their consent. In this context, the individual circumstances, physical abilities, the ability to consent/ anticipate the consequences and the conception on the safety of the person can be determined. This ensures that solutions and achievements by the system and other alternatives can be explained, proposed and discussed. The user is supported to deal with the provision and use of his personal data depending on his competence and values. Based on the provided information they have the option to configure settings and access authorization and are free to choose and use the provided system functionality. The transparency is an important aspect when dealing with personal data, so the users will be able to inquire their saved personal data at any time using the system.

To provide the user with adequate comprehension about the system functionality, data acquisition, collection and processing, explanations and information are presented appropriate regarding the target audience of elderly people. Depending on the user's preference more detailed information, system settings and permissions can be accessed. In combination with pre-defined profiles to grant access authorization and access permissions for involved persons a comprehensible system knowledge can be established and a demand-oriented configuration and usage can be provided.

Because the system contains an electronic medication plan, it was discussed whether there should be a review of the interactions and contraindications when importing a new medication plan, and a warning should be displayed if the result is positive. As it is possible that this functionality could cause further non-targeted actions by the user, e.g. stop taking the drug without seeking medical advice. As this could be contrary to the actual benefit of the medication plan. The function is not implemented until there is found an acceptable solution that does ensure harmless use.

4. Discussion

In this work we included the model MEESTAR in a structured and systematic methodology to identify ethical and moral conflicts successfully in the process of concepting and implementing a patient-centered system within the PIQ project. The given dimension, different points of view, fundamental ethical and moral conflicts and associated questions were used to critically reflect the functionality and possible impacts on those who use and integrate the system in their daily life. The inclusion and discussion of ethical aspects led to encourage awareness and consideration of ethical assessment in the ongoing designing and implementing of the system.

By extending and specifying ethical questions specific measures and solutions could be developed that are further enhanced and monitored in an iterative process. We derived concrete solutions for specific scenarios and their related ethical conflicts based on the presented proceeding (sec 2.2). The specification is necessary because the pre-defined ethical conflicts and guidelines provided by MEESTAR are partially to general and had to be extended to provide a complete consideration of the functionality of the system.

However, results presented in this work are based on expert knowledge and the developed process only. In future work, we will evaluate consequences and effects that arise from the developed solutions with appropriate quantitative and qualitative methods: we will make use of focus groups to include the potential system users and measure their accordance with the identified scenarios, and ethical conflicts presented in this work using the developed process and expert knowledge.

References

[1] VIEHWEGER, Axel ; KARMANN, Alexander ; UHLMANN, Michael: Selbstständiges Wohnen bis ins hohe Alter : Eine volkswirtschaftliche Analyse. Dresden : VSWG, 2015

[2] MARING, Matthias [Hrsg.: Vom Praktisch-Werden der Ethik in interdisziplinärer Sicht : Ansätze und Beispiele der Institutionalisierung, Konkretisierung und Implementierung der Ethik. s.l. : KIT Scientific Publishing, 2015, pp. 299-341.

[3] MANZESCHKE, Arne ; WEBER, Karsten ; ROTHER, Elisabeth ; FANGERAU, Heiner: Ethical Questions in the Area of Age Appropriate Assisting Systems. Ludwigsfelde : Druckerei Thiel Gruppe, 2015

[4] FLEMMING, Daniel ; GIEHOFF, Carsten ; HÜBNER, Ursula: Entwicklung eines Standards für den elektronischen Pflegebericht auf Basis der HL7 CDA Release 2. In: Pflegewissenschaft (2008), Nr. 12, pp. 676–682.

[5] MÜLLER-MIELITZ, Stefan ; LUX, Thomas: E-Health-Ökonomie. Wiesbaden : Springer Fachmedien Wiesbaden, 2017, pp. 821-847.

[6] MÖLLER, Horst ; ALY, Amin-Farid: Definitionen zu Pharmakovigilanz und Arzneimitteltherapiesicherheit (AMTS). In: Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen 106 (2012), Nr. 10, pp. 709–711.

[7] KUSKE, S. ; LESSING, C. ; LUX, R. ; SCHMITZ, A. ; SCHRAPPE, M.: Patientensicherheitsindikatoren zur Arzneimitteltherapiesicherheit (AMTS-PSI) : Internationaler Status, Übertragbarkeit und Validierung*. In: Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)) 74 (2012), Nr. 2, pp. 79–86.

[8] PFANNSTIEL, Mario Alexander (Hrsg.); KRAMMER, Sandra (Hrsg.); SWOBODA, Walter (Hrsg.): Digitale Transformation von Dienstleistungen im Gesundheitswesen III : Impulse für die Pflegepraxis. Wiesbaden : Springer Fachmedien Wiesbaden, 2017, pp. 71-87.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-286

Proof of Concept of a Partial Weight-Bearing Supporting Real-Time Feedback System

Andrés I. TKACHENKO-BRIL[a, 1], Harald JAGOŠ[b], Veronika DAVID[c], Katharina PILS[d], Jakob GAUDERNAK[d] and Dietmar RAFOLT[a]
[a]*Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria*
[b]*Raltec - Research group for assisted living technologies, Austria*
[c]*Department for Biomedical, Health & Sports Engineering, UAS Technikum Wien, Austria*
[d]*Institute of Physical Medicine and Rehabilitation, Sophienspital (Vienna), Austria*

Abstract. Background: Patient compliance with lower limb partial weight bearing (PWB) instructions during post-surgical early mobilization is often low and many are unable to adhere to the prescribed limits without the implementation of concurrent biofeedback. Objectives: A real-time feedback system based on eSHOE instrumented insoles was tested in order to preliminarily quantify its efficiency at improving geriatric patients' compliance. Methods: In order to gain a proof of concept, measurements with one patient after a hip fracture were carried out. The compliance with the prescribed load restriction was measured on four measurement dates, first without and then with the feedback. The number of correctly loaded steps (NCS), the mean peak load (MPL) and the maximal load (ML) were considered. Results: Preliminary results of one patient show that NCS was nearly doubled and the MPL was reduced to acceptable limits, while the ML was reduced on three of the four days. Conclusion: The results indicate that the developed system is easily implementable into the rehabilitative routine and has a positive effect on PWB performance of geriatric subjects while walking.

Keywords. Early ambulation, weight-bearing, patient compliance, feedback.

1. Introduction

Early mobilization after lower body injuries has become the norm lately due to the increased awareness of the importance of muscle strength and bone density preservation [1]. In order to minimize the risks and optimize the benefits of early mobilization, a partial weight-bearing (PWB) protocol of about 10-50% of the patient's bodyweight is prescribed. The critical downside to this approach is that the compliance with the weight limitations of even young and healthy subjects has been shown to be poor [2]. Elderly subjects, whose diminished strength, diverse comorbidities as well as limited exterio- and proprioceptive capacities impede their weight bearing assessment ability, perform even worse [3]. At the same time, geriatric patients are operated most often on the hip and in the area of the femoral neck [4]. Pathologies like upper body weakness and disorders of gait or cognition can make it almost impossible to understand to which

[1]Corresponding author: Andrés I. Tkachenko-Bril. Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria. Andres.tkachenkobril@meduniwien.ac.at

degree geriatric patients follow the prescriptions [2]. Literature shows that 90% of the patients end up overloading the limb as soon as they leave the scale on which they are shown how the prescribed load feels, especially while climbing stairs [5]. Without any sensors or feedback mechanisms, the actual loads acting in the lower body can be neither measured nor monitored.

Improving geriatric patients' PWB performance is difficult due to the current standard procedures. The doctor or surgeon usually recommends the PWB limit, but it is in the rehabilitation centers where patients learn how to limit the peak loads they apply onto their injured limb. Physical therapists use different methods that vary from one center to another, as there is no golden standard [2]. The outcome of the teaching process depends strongly on the therapist's experience. However, several authors have stated, that the most commonly used techniques to teach PWB, namely weight scales and verbal cues, are ineffective as is it difficult for the patients to follow the indications [6]. The aforementioned techniques' efficiency has been found to improve with the implementation of a concurrent feedback [7]. Additionally, the retention or learning effect of the patients that are trained with such devices has been shown to be superior [8]. However, feedback systems are finally not available to patients in a physical rehabilitation context due to their elevated price, complexity or low acceptability by the patients [2].

In an attempt to improve the current PWB application methodology in rehabilitation centers, a cheap, simple and unobtrusive system, called "eSHOE", was developed in previous research projects [9-12]. In the course of eSHOE's further development, the biofeedback device was tested in a clinical context in a proof-of-concept approach to obtain a preliminary insight of the feasibility of its incorporation into PWB teaching and practice routines of rehabilitation centers. The main objectives were to obtain a first impression and feedback from the physical therapists and patients and to quantify the effect that the system had on the patient's performance.

2. Methods

In order to test the concept of the novel biofeedback system, a series of measurements were conducted with one patient during her inpatient stay at the institute of physical medicine and rehabilitation of the Sophienspital hospital in Vienna. The patient was instructed to wear the eSHOE insoles while practicing PWB in the context of the regular rehabilitative process on four measuring dates. The first three measurement days (MD) took place on three days as close together as possible. The last MD was conducted shortly before the discharge of the patient from the hospital. Following the gold standard when teaching partial weight bearing mobilization, a force plate was used on the first MD to teach and practice the feel of the prescribed load, which then the patient was instructed to emulate while walking. During this procedure the threshold value for the prescribed load was also measured via eSHOE and set on the application. On all measurement dates, the patient walked twice along a 15 m long path with a walking aid and two conditions were measured. In the first one, motion parameters were recorded with the biofeedback system but no feedback was provided (NF). In the other, motion parameters were recorded and the patient was provided with concurrent feedback. The main parameters that were extracted from the collected data were the number of correct steps as a percentage of the total steps and each step's peak load as a percentage of the patient's

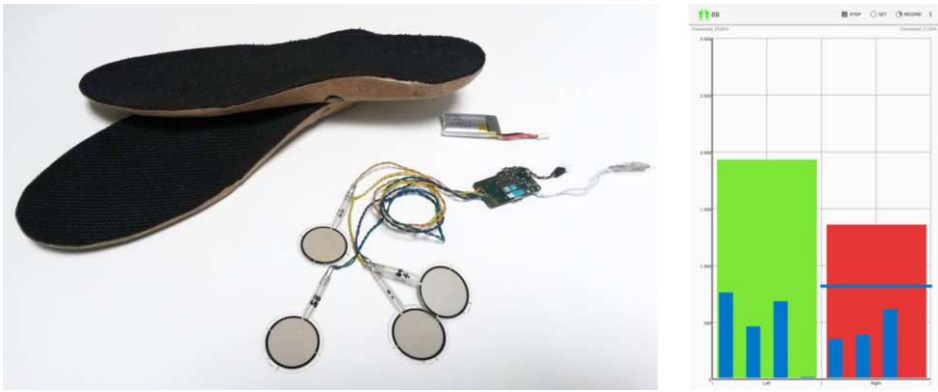


Figure 1 Left: The multimodal biometric measurement system for mobile gait analysis and therapy monitoring, eSHOE. The embedded electronic components can be seen besides the insoles. Right: Screenshot of the smartphone biofeedback application while the load applied on the right foot is monitored, the load being applied has exceeded the predefined threshold value (blue horizontal line).

bodyweight (BW) during both conditions. A step was deemed as 'correct' if its peak load remained in the ±10% BW range around the prescribed PWB limit.

The subject referenced in this article gave her informed consent prior to testing. All calculations were performed in MATLAB (The Mathworks, Massachusetts, USA).

2.1. eSHOE instrumented insoles

The biofeedback system is based on the multimodal biometric measurement system for mobile gait analysis and therapy monitoring eSHOE system (Figure 1, left).

The relevant sensors of the eSHOE system in the scope of this paper are the four force sensitive resistors (FSR), which measure the pressure under the heel, the metatarsal heads I and V, and the big toe. These sensors offer a sensitivity of 11 g/LSB, as the sampling resolution of the microcontroller's ADC module is 10 Bit. All sensors' data is gathered at 200 Hz and can be saved or sent to a compatible device over Bluetooth.

2.2. Live feedback Android application

The Android application (Figure 1, right) provides the user with a simple interface and three feedback modes (visual, acoustic, haptic). It was designed to cover the physical therapists' expectations that were identified previously [13]. Four different feedback modes are implemented with the intention of satisfying the heterogeneous needs of the patients. Visual, acoustic and/or haptic stimuli can be set to provide the user with feedback [14]. During the calibration procedure, residual loads are first canceled out while the patient is sitting. Then, the threshold value used to trigger the feedback response is set while the patient applies the prescribed load with the affected limb on a weight scale or force plate (Figure 2, left). The value is set 5% lower than the mean of the last 10 received values in order to account for the overshoot phenomenon described by Warren and Lehmann [15]. Additionally, an "acceptable load" range can be set around the threshold value, usually ±10% of the patient's BW.

Figure 2: (a) The calibration process of the biofeedback system using a force plate. (b) Patient becoming visual feedback during a measurement.

3. Results

The patient, a 74-year-old female, was prescribed a 50% bodyweight (BW) limit for four weeks after a medial femur neck fracture, which was treated with the implantation of a sliding hip screw. The patient had already been introduced to the PWB. The first three measurements were carried out in the first two weeks and the last one was conducted two days before the patient was discharged. The mean load measured difference with the prescribed load in BW percentage points is depicted in Figure 3. Approximately 30 steps were recorded and analyzed under each of the conditions on each of the four measurement days. The subject preferred to receive an acoustic cue whenever the load was too high. On the other hand, the therapist was able to see the patient's performance as a visual feedback in real-time and could advise the patient accordingly. The system was calibrated at the beginning of each of the measurement days.

The obtained data is shown in Figure 3. On MD1, the median mean load applied by the patient right limb after the training with the force plate was -10.4 % BW. Under the FB condition, the load increased to 9.2 % BW; whereas the maximal load went from 8.0 % BW to 26.1 % BW. The percentage of correctly loaded steps rose from 43.2 % to 53.6 %.

On MD2, the median mean load was reduced from 16.7 % BW to 6.1 % BW. The maximal recorded load was 34.7 % BW above the prescribed load without the feedback and went down to 25.1 % BW. The number of correctly loaded steps grew from 25.9 % to 48.3 % of the total steps.

On MD3, the median load under NF was 20.3 % BW. It was reduced to 4.8 % BW above the prescribed load with the implementation of feedback. The maximal load was reduced from 30.5 % BW to 20.1 % BW. The proportion of correctly loaded steps increased from 5.7% to 73.2 %.

On MD4, the median mean load went from 18.9 % BW to 9.1 % BW. The maximal recorded load under the NF condition was 32 % BW, and it fell to 20.6 % BW under the FB condition. All steps were overloaded without feedback, whereas 58 % of the total steps were loaded correctly under the feedback condition.

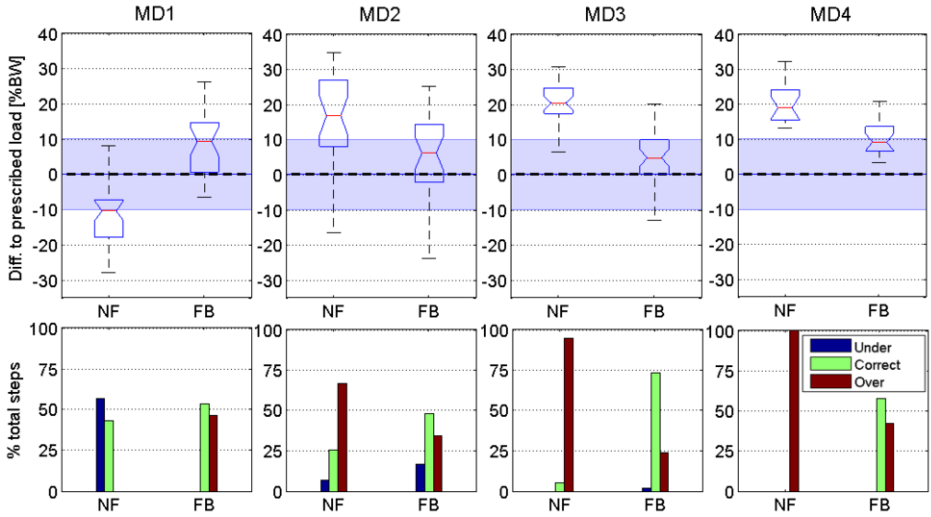


Figure 3: (Upper plot) Peak step loads recorded under the no-feedback (NF) and feedback (FB) conditions on each of the measurement days (MD) as differences to the prescribed load. The blue area depicts the 'acceptable load range'. (Lower plot) Percentage of steps that was below (blue), above (red) or met (green) the prescribed load.

4. Discussion

The results show that the patient's compliance with the PWB load restriction was low, as it was expected from the results observed in the literature reviewed. The implementation of the biofeedback system improved her compliance and confidence, but a learning effect over time was not observed. These findings are consistent with the observations done in previous studies. With feedback, patients are mostly able to adhere to the prescribed loads, but this improvement disappears shortly afterwards [14]. With some further development of the battery duration of the developed system, patients could be provided with haptic feedback at all times during their inpatient stay to improve the overall compliance.

The implementation of the biofeedback system was valued in a positive way by the patient and the medical personnel. Its implementation did not disturb the patient's gait, as reported by all parties, and the feedback function worked as expected. The patient expressed that the system made her feel more confident. On the first MD, the patient was afraid to load too much weight after the short training procedure, which resulted in an unnatural gait and an evident under-loading of the limb. The feedback and assurance from the physical therapist that more load was tolerated improved both issues. More confidence was observed on the next three measuring dates, when the median step load grew above the acceptable range under NF. The feedback via the application helped the patient on all occasions to shift the difference between the measured median peak load and the prescribed load into the acceptable interval of 0±10 % BW. The maximum load recorded was also reduced on all but the first MD. The amount of correctly loaded steps as a percentage of the total steps increased under the FB condition on all MDs.

The physical therapist indicated that the visual feedback was of great value, as it is otherwise impossible to tell how the patient is performing while walking. The cues that

the therapist was able to provide to the patient were more accurate with the real-time information provided by the biofeedback system. Additional pathologies, that are invisible to the naked eye, could be unveiled, e.g. non-loading of the heel because of fear of overloading the affected leg on the first MD. According to the physical therapist the configuration and implementation of the system was not complicated or demanding. The possibility to view the patient's performance in real time as well as to record the measurements to quantify a progression during the rehabilitative process were deemed as important characteristics.

In conclusion, the limited significance of the presented results is evident as the system was only tested on one patient. However, the biofeedback device has proven to be implementable in the clinical context of partial weight bearing and was positively evaluated by both the patient and the therapist. The positive effect of the system has been preliminarily quantified and verified and has suggested that further development and assessment via larger clinical studies of the system is desirable.

The next logical step is to conduct a larger clinical trial, in which the subject cohort size would allow for a comparison with a control group after validating the biofeedback system against a plantar pressure measuring gold standard.

5. Acknowledgemnents

The study was conducted as part of the project "REHABitation", which is funded by the Austrian Research Promotion Agency (FFG) within the funding program COIN.

6. References

[1] L. B. Oldmeadow, E. R. Edwards, L. A. Kimmel, E. Kipen, V. J. Robertson, and M. J. Bailey, "No rest for the wounded: early ambulation after hip surgery accelerates recovery.," ANZ J. Surg., vol. 76, no. 7, pp. 607–611, Jul. 2006.

[2] J. W. Hustedt, D. J. Blizzard, M. R. Baumgaertner, M. P. Leslie, and J. N. Grauer, "Current advances in training orthopaedic patients to comply with partial weight-bearing instructions," Yale J. Biol. Med., vol. 85, no. 1, pp. 119–125, 2012.

[3] J. W. Hustedt, D. J. Blizzard, M. R. Baumgaertner, M. P. Leslie, and J. N. Grauer, "Effect of age on partial weight-bearing training.," Orthopedics, vol. 35, no. 7, pp. e1061-7, Jul. 2012.

[4] I. B. Menzies, D. A. Mendelson, S. L. Kates, and S. M. Friedman, "The impact of comorbidity on perioperative outcomes of hip fractures in a geriatric fracture model.," Geriatr. Orthop. Surg. Rehabil., vol. 3, no. 3, pp. 129–34, Sep. 2012.

[5] C. J. Winstein, P. S. Pohl, C. Cardinale, a Green, L. Scholtz, and C. S. Waters, "Learning a partial-weight-bearing skill: effectiveness of two forms of feedback.," Phys. Ther., vol. 76, no. 9, pp. 985–993, 1996.

[6] H. L. Hurkmans, J. B. Bussmann, E. Benda, J. A. Haisma, J. A. Verhaar, and H. J. Stam, "Predictors of partial weight-bearing performance after total hip arthroplasty," J. Rehabil. Med., vol. 42, no. 1, pp. 42–48, 2010.

[7] H. L. Hurkmans, J. B. Bussmann, E. Benda, J. A. Verhaar, and H. J. Stam, "Effectiveness of audio feedback for partial weight-bearing in and outside the hospital: A randomized controlled trial," Arch. Phys. Med. Rehabil., vol. 93, no. 4, pp. 565–570, 2012.

[8] Z. Pataky, D. De León Rodriguez, A. Golay, M. Assal, J.-P. Assal, and C.-A. Hauert, "Biofeedback training for partial weight bearing in patients after total hip arthroplasty.," Arch. Phys. Med. Rehabil., vol. 90, no. 8, pp. 1435–8, Aug. 2009.

[9] H. Jagos et al., "Mobile gait analysis via eSHOEs instrumented shoe insoles: a pilot study for validation against the gold standard GAITRite ®," J. Med. Eng. Technol., vol. 41, no. 5, pp. 375–386, Jul. 2017.

[10] H. Jagos et al., "Determination of Gait Parameters from the Wearable Motion Analysis System eSHOE.," Biomed. Tech. (Berl)., vol. 58, pp. 4–5, 2013.

[11] H. Jagos, V. David, M. Reichel, and S. Kotzian, "Tele-Monitoring of the Rehabilitation Progress in Stroke Patients," no. Figure 1, pp. 311–313, 2015.
[12] H. Jagos, J. Oberzaucher, M. Reichel, W. L. Zagler, and W. Hlauschek, "A multimodal approach for insole motion measurement and analysis," Procedia Eng., vol. 2, no. 2, pp. 3103–3108, 2010.
[13] A. I. Tkachenko-Bril, V. David, M. Scherer, H. Jagos, P. Kafka, and A. Sabo, "Development of a Wearable Live-feedback System to Support Partial Weight-bearing while Recovering from Lower Extremity Injuries," Procedia Eng., vol. 147, pp. 157–162, 2016.
[14] H. L. Hurkmans, J. B. Bussmann, R. W. Selles, and E. Benda, "The Difference Between Actual and Prescribed Weight Bearing of Total Hip Patients With a Trochanteric Osteotomy : Long-Term Vertical Force Measurements Inside and Outside the Hospital," vol. 88, no. February, pp. 200–206, 2007.
[15] C. G. Warren and J. F. Lehmann, "Training procedures and biofeedback methods to achieve controlled partial weight bearing: an assessment.," Arch. Phys. Med. Rehabil., vol. 56, no. 10, pp. 449–55, Oct. 1975.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-293

Towards Phenotyping of Clinical Trial Eligibility Criteria

Matthias LÖBE [a,1], Sebastian STÄUBERT [a,b,1], Colleen GOLDBERG [a], Ivonne HAFFNER [c] and Alfred WINTER [a]
[a]*Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Universität Leipzig, Germany*
[b]*Clinical Trial Centre Leipzig (ZKSL), Universität Leipzig, Germany*
[c] *University Cancer Center Leipzig, Germany*

Abstract. Background: Medical plaintext documents contain important facts about patients, but they are rarely available for structured queries. The provision of structured information from natural language texts in addition to the existing structured data can significantly speed up the search for fulfilled inclusion criteria and thus improve the recruitment rate. Objectives: This work is aimed at supporting clinical trial recruitment with text mining techniques to identify suitable subjects in hospitals. Method: Based on the inclusion/exclusion criteria of 5 sample studies and a text corpus consisting of 212 doctor's letters and medical follow-up documentation from a university cancer center, a prototype was developed and technically evaluated using NLP procedures (UIMA) for the extraction of facts from medical free texts. Results: It was found that although the extracted entities are not always correct (precision between 23% and 96%), they provide a decisive indication as to which patient file should be read preferentially. Conclusion: The prototype presented here demonstrates the technical feasibility. In order to find available, lucrative phenotypes, an in-depth evaluation is required.

Keywords. Text Mining, Clinical Trials, Recruitment, Phenotyping, NLP, Apache UIMA, cTAKES

1. Introduction

A common problem in clinical research is slow and meager recruitment of subjects for clinical trials. There are several reasons for this. If the attending physician is also involved as an investigator in an academic trial, the risk of overestimating the number of patients can simply be a matter of enthusiasm. The researchers overestimate the number of eligible patients because they only see the underlying disease, but many subjects are unsuitable for other reasons. Furthermore, the study protocols are becoming more complex, resulting in an increasing number of inclusion and exclusion critiques and more complicated interventions. On the other hand, a large number of patients do not want to take part in clinical trials or will drop out later because the benefit is not perceived. Ultimately, patients are often not addressed at the right time, e. g. if the attending physician does not know the study during the hospital stay [1,2]. As a result, clinical studies take longer, are more expensive, compromise on meaningfulness (lower sample sizes) or are completely discontinued, resulting in the delayed availability of new drugs

[1] Corresponding Authors: Matthias Löbe and Sebastian Stäubert (equally contributed), Institute for Medical Informatics, Statistics and Epidemiology, Universität Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany; Email: matthias.loebe@imise.uni-leipzig.de, sebastian.staeubert@imise.uni-leipzig.de

and therapies. However, computers can support the investigator's work in various ways [3]. To include (or exclude) subjects for a clinical trial, a set of eligibility criteria is defined. Inclusion criteria often include factors that indicate the clinical picture studied in the study, such as demographic data, diagnoses or laboratory values. Exclusion criteria include circumstances that would adversely affect study participation, such as pregnancy, use of addictive substances, pre-existing conditions, or concomitant medication. Not all inclusion/exclusion criteria can be queried automatically on the available data in the EMR, because they are either not recorded at all or not completely (for example if they are not relevant for claims to the health insurance). In this case, cohorts of patients are often formed who meet the criteria that can be queried automatically and study nurses read the patient's file in search of further fulfilled criteria. Manual reading is time-consuming and therefore often just as incomplete. If not all the desired criteria can be taught in this way, they are collected in the context of a pre-screening visit, which is outside the focus of this paper. Our approach is to relieve the study nurse from the burden of reading large amounts of patient records by using natural language processing (NLP) techniques to identify certain characteristics in the patient file. Thus, only files of selected patients have to be investigated. The process of mining patient data to classify subjects into distinct classes by using a set of observable characteristics of an individual as a classifier is often called *phenotyping* [4-6]. Therefore, eligibility criteria can be regarded as phenotypes.

2. Objectives

The aim of the project was to create a prototype tool to support patient recruitment for clinical trials. The most time-consuming process is the search for suitable candidates from the entirety of all patients (screening). Currently, study nurses still have to read large parts of the electronic patient record of all eligible subjects in order to investigate the eligibility or reasons for not being eligible. This work can be facilitated by a computer-assisted system that identifies key words or important phrases and highlights them in the right context. To this end, inclusion and exclusion criteria are to be formalized and medical free text documents are to be analyzed by means of an NLP pipeline. Thus, the study nurses would prefer to read such files in which the relevant phenotypes (e.g. diagnoses or procedures) are at least mentioned, even if the algorithmic recognition of the circumstances does not work perfectly.

3. Methods

The following procedure was chosen for the implementation of the objectives:

1. Creation of a corpus of medical plaintext documents
2. Selection of a set of suitable studies and compilation of inclusion and exclusion criteria as target phenotypes
3. Research of suitable vocabularies to reflect the terminology used in patient records
4. Construction of an NLP pipeline und execution
5. Determination of statistical metrics und evaluation of results

Table 1. Selected studies from the university cancer center

Study name	Study type	Condition	NCT number[1]
ADAPT	phase 2, 3	Breast Cancer	NCT01779206
MATEO	phase 2	Metastatic, Esophagogastric Adenocarcinoma	NCT02128243
VARIANZ	observational	Esophageal Neoplasms, Stomach Neoplasms	NCT02305043
MONALeesa	phase 3	Advanced, Metastatic Breast Cancer	NCT02278120
OLYMPIA	phase 3	Breast Cancer	NCT02032823

The text corpus we used consists of 212 plaintext documents from 101 patients of the Leipzig University Cancer Center (UCCL). The documents were of two types: medical progress documentation and doctor's letters. The planned sample size had to be significantly reduced both in terms of volume and the number of supported document classes, because it turned out that no automatic export of all documents from a patient's record from the hospital information system is currently possible, so that the files had to be extracted manually, which of course was an additional effort. A reference corpus of German medical records for research is currently not available.

Five oncological studies were selected (cf. Table 1), all of which were actively conducted in 2015 in the UCCL and where the recruitment rate was below expectations. The free text inclusion and exclusion criteria (cf. "Eligibility Criteria" in the "Tabular View" of the 5 selected studies in the respective entry at clinicaltrials.gov) were analyzed with regard to medical domain, explicit definitions and range of values. In order to keep the prototype manageable, we focused on the two most common domains, diagnoses and laboratory values.

We were looking for suitable vocabularies that contained concepts with the textual labels corresponding to the phenotypes to be found in the medical documents. Diagnoses in Germany are coded with ICD-10-GM (German modification). ICD-10 provides only a single textual label for each diagnosis code. That's why we also used Alpha-ID. Alpha-ID is a thesaurus of colloquial German-language disease descriptions with a reference to ICD-10 codes. In addition, we wanted to distinguish between mentioning a diagnosis and excluding it. For this purpose, we have compiled a list of words for words with negating meaning based on the Duden Synonym dictionary [7]. Laboratory values usually consist of a name, a value and a measurement unit. Unfortunately, there is no uniform standard for the name of the analytes in Germany, so that a separate list of different sources has been compiled. For units, the coding system Unified Code for Units of Measure (UCUM) was used.

The next step was to plan and implement the NLP pipeline. The pipeline is based on Apache UIMA [8] and Apache cTAKES [9]. Both are open source software. The UIMA framework provides components, interfaces, data representation and patterns for the construction of such pipelines. cTAKES is an implementation of a UIMA pipeline with special components (so-called annotators), which can process clinical documents both rule-based and using machine learning methods. cTAKES supports XML, plaintext and Clinical Document Architecture (CDA) as input formats and is trained on an English clinical documentation body.

A pipeline is usually divided into several modules that are executed in succession. This approach enables the reusability of modules and flexibility in the sequence of sub-steps. Our pipeline consists of 3 sections: pre-processing, document processing and the output unit (see Figure 1). In the preprocessing section, full-text documents are read from

[1] ClinicalTrials.gov Identifier

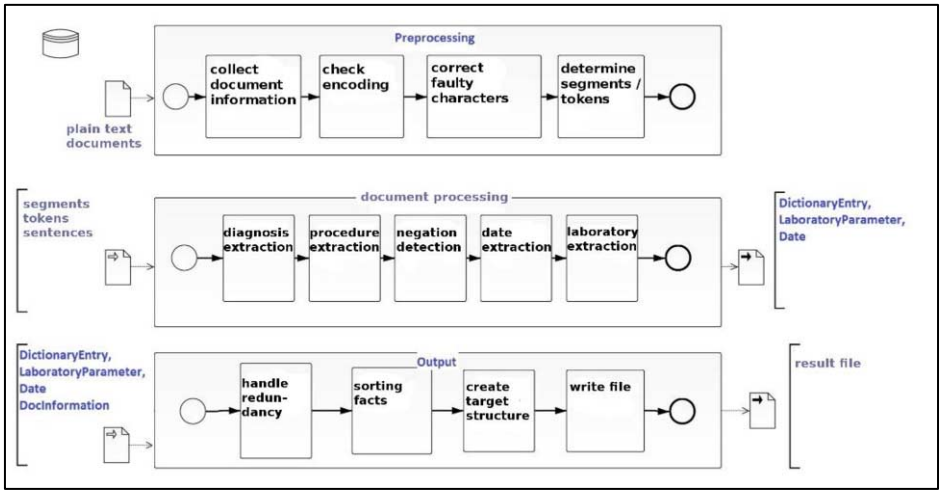


Figure 1. NLP Pipeline with its preprocessing, document processing and output units

a directory and converted from Microsoft XPS (doctor's letter) and Microsoft DOC (follow-up documentation) to plaintext. Then the encoding is checked and segments, sentences and tokens are recognized in the text. After this step, segments, records and tokens exist as objects and form the input for processing. The processing section begins with the annotation of diagnoses. Negations are searched for in the context of detected diagnoses. The negation itself does not form its own annotation type, but only a feature of the diagnosis type. Laboratory parameters are then extracted with value and unit. In the output unit, all collected information is cleaned up of redundancies and assigned to the corresponding patient. The pipeline is completed by writing the results to a CSV file. The NLP pipeline prototype is available via Leipzig Health Atlas [10].

4. Results

The available document body was divided evenly into a training set and an evaluation set. 18 documents were empty and were previously removed. Both sets of documents contained approximately the same number of patients, of doctor's letter and follow-up documentation. The total length of the documents (72,309 vs. 79,295 tokens) was similar in both sets too. Due to data protection restrictions, it was not possible to create a gold standard using annotation by external experts. The laboratory values were annotated by the authors themselves. The list of negative words has also been extended to include other expressions. The evaluation of the results of the NLP pipeline on the evaluation set was also carried out manually.

We used to following performance metrics to measure quality and completeness. All metrics have a value range from 0 to 1, the latter being the perfect value.

$$Precision = \frac{\text{TruePositives}}{\text{TruePositives+FalsePositives}} \quad (1)$$

$$Recall = \frac{\text{TruePositives}}{\text{TruePositives+FalseNegatives}} \quad (2)$$

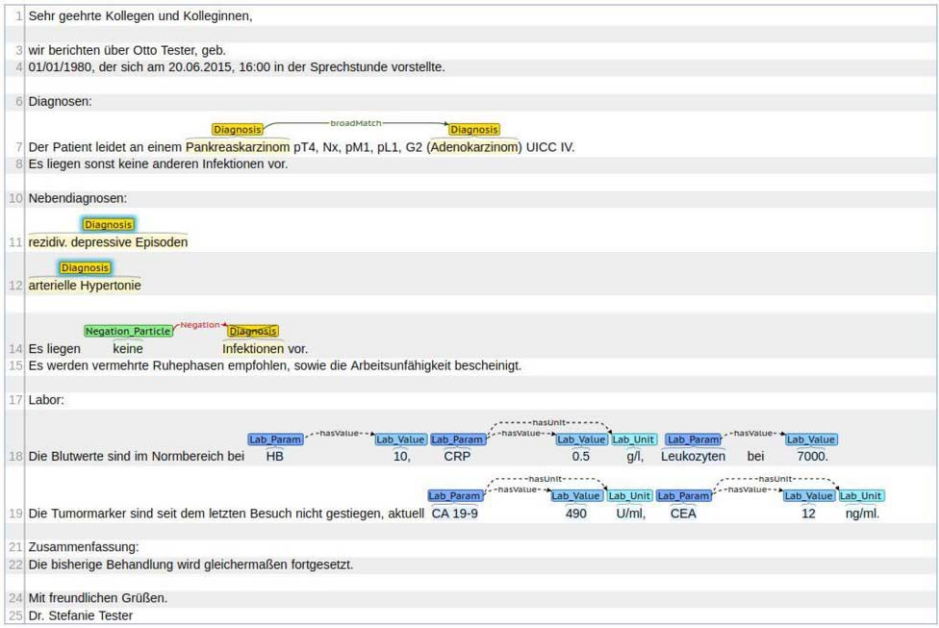


Figure 2. Example of a manual annotation with the web-based text annotation tool brat [11]

$$Accuracy = \frac{\text{TruePositives}+\text{TrueNegatives}}{\text{TruePositives}+\text{TrueNegatives}+\text{FalsePositives}+\text{FalseNegatives}} \quad (3)$$

$$F\ score = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Each of the diagnoses found by the pipeline was compared manually with the documents in the evaluation set. Each diagnosis found was divided into the group "correct annotation" or "incorrect annotation" according to the evaluation rules. A total of 1,105 annotations were found. Of these, 1.071 were correct and 34 were false. This results in a positive predictive value (precision) of 1.071/1.105 = 96.9%. The missing diagnoses were also determined. With 238 missing annotations, the sensitivity (recall) is 1.071/1.309 = 81.8%. The F score is 0,89.

Two lexicons with analyte names and units of measurement are available for the annotation of laboratory analytes. Two user defined parameters have been created in the annotator. These determine the maximum number of tokens between the analyte name and value (x) or analyte name and unit (y). The best results were obtained for x=3 and y=5. With 127 annotations, 64 were wrongly detected. This results in a positive predictive value (precision) of 63/127 = 49.6%. Of the 100 laboratory values in the evaluation set, 63 were correctly annotated, so the sensitivity (recall) is 63%. The F-score is 0.56.

For diagnoses, the system checks whether a statement is positive or negative. This is a binary qualifier. The measure that can be used here to evaluate the outcome of the negation is accuracy. For this purpose, each diagnosis of the evaluation corpus is divided into one of the categories "positive diagnosis" or "negative diagnosis". Negative statements are the TruePositives and positive statements are the TrueNegatives. The

window size to the maximum distance of the negation word from the diagnosis was 3. The accuracy was (35+1.184)/ (35+12+41+1.184) = 95.8%.

5. Discussion and Outlook

The work presented here shows a prototypical but functional pipeline for extracting diagnoses and laboratory findings from clinical documents. The study recruitment use case is particularly suitable for NLP support. Some typical metrics in statistical tests (sensitivity, negative predictive value) play a minor role here. Given the low prevalence of the "eligibility phenotype" of a typical clinical trial in the whole population of patients of a hospital department and the rapid and resource-friendly execution of computer-based analyses compared to manual reading, worthwhile patient records can be selected much more precisely.

The good performance in terms of sensitivity in the detection of diagnoses is explained by the quality of the available dictionary (Alpha-ID) and could be further enhanced by the inclusion of abbreviations [12]. Laboratory values are much more difficult to recognize, not only because of the missing dictionary, but also because of their complex expression (name + value + unit of measurement), which is complicated by typos, abbreviations or normal ranges. The good result for negation detection is mainly explained by the high number of TrueNegatives.

The current approach has a number of limitations. Firstly, only a limited number of documents could be made available, originating from only two different document classes. The 5 clinical trials all originate from the field of oncology and the inclusion and exclusion criteria are not representative. Furthermore, only diagnoses and laboratory values were examined. Diagnoses and laboratory values are typically part of the structured data available in the HIS and could be queried automatically. However, it is not uncommon for certain diagnoses not to be coded for billing reasons, for patients to move from other facilities to a cancer center, or for technical and administrative barriers to programmatic access to existing data in other departments. Other types of medical data would benefit much more from an NLP approach, such as the search for allergies, implants or certain lifestyle-related factors such as smoking or alcohol consumption.

The evaluation was not carried out by clinical personnel, but by computer scientists. To date, no tests has been carried out as to whether the approach accelerates the screening process or improves the recruitment rate by making better use of resources.

In the future, we plan to extend the NLP pipeline to include other kinds of patient data, for instance vital signs, medications, procedures and lifestyle. Another focus is the use of the resulting structured data for retrieval in the research data warehouse of the university hospital (i2b2, tranSMART). In general, more work is needed in the field of German-language text analysis. While there is a large selection of tools, document catalogs and research initiatives for English [13], there are no comparable structures for German.

Finally, electronic health records-driven phenotyping will be a decisive topic for the data integration centers of the German Medical Informatics Initiative, a new 150 million euros funded program to make data from health care available for research across sites [14]. The University of Leipzig is part of the SMITH consortium [15] of the MI Initiative. A methodical use case in SMITH is the Phenotype Pipeline. Among other things, NLP components are also planned here, which follow a similar approach to the one presented here but are much more extensive. In view of the comparability of phenotypes across site

boundaries [16], storage and availability in central open Metadata Repositories (MDR) is recommended.

Acknowledgements

The work of this project was part of the master thesis of Colleen Goldberg [17] and supported by the German Ministry of Education and Research, funding reference: 01KN1102.

References

[1] Fletcher B, Gheorghe A, Moore D, Wilson S, Damery S. Improving the recruitment activity of clinicians in randomised controlled trials: A systematic review. BMJ open 2012;2(1):e000496.
[2] Thoma A, Farrokhyar F, McKnight L, Bhandari M. Practical tips for surgical research: How to optimize patient recruitment. Canadian journal of surgery. Journal canadien de chirurgie 2010;53(3):205–10.
[3] Köpcke F, Prokosch H-U. Employing Computers for the Recruitment into Clinical Trials: A Comprehensive Systematic Review. J. Med. Internet Res. 2014;16(7):e161.
[4] Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. Journal of the American Medical Informatics Association JAMIA 2013;20(1):117–21.
[5] Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: Challenges, recent advances, and perspectives. Journal of the American Medical Informatics Association JAMIA 2013;20(e2):e206-11.
[6] Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: Towards better research applications and clinical care. Nature reviews. Genetics 2012;13(6):395–405.
[7] Duden K, Wermke M. Der Duden in zwölf Bänden: Das Standardwerk zur deutschen Sprache. Mannheim: Dudenverlag; 2002-2008.
[8] Ferrucci D, Lally A. UIMA: An architectural approach to unstructured information processing in the corporate research environment. Nat. Lang. Eng. 2004;10(3-4):327–48.
[9] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. Journal of the American Medical Informatics Association JAMIA 2010;17(5):507–13.
[10] Goldberg C, Löbe M, Stäubert S. NLP4CR Pipeline; Available from: https://www.health-atlas.de/en/method/nlp4cr-pipeline.
[11] Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J'i. BRAT: A Web-based Tool for NLP-assisted Text Annotation. In: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics; 2012, p. 102–107.
[12] Long W. Extracting diagnoses from discharge summaries. AMIA Annu. Symp. Proc. 2005:470–4.
[13] Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association JAMIA 2011;18(5):552–6.
[14] TMF e.V. Medical Informatics Initiative Germany. [February 11, 2018]; Available from: http://www.medizininformatik-initiative.de/en/about-initiative.
[15] Löffler M, Scherag A, Marx G. Smart Medical Information Technology for Healthcare (SMITH). [November 12, 2017]; Available from: http://www.smith.care/.
[16] Pathak J, Wang J, Kashyap S, Basford M, Li R, Masys DR et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: The eMERGE Network experience. Journal of the American Medical Informatics Association JAMIA 2011;18(4):376–86.
[17] Goldberg C. Unterstützung des Rekrutierungsprozesses für klinisches Studien durch Natural Language Processing auf klinischen Dokumenten. Master Thesis. Leipzig; 2016.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-300

Improving Patient Safety by Reusing Clinical Routine Data – An Expert Survey on Patient Safety Indicators

Michael SCHALLER[a,1], Werner O. HACKL[a] and Elske AMMENWERTH[a]
[a]*Institute of Medical Informatics, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*

Abstract. Background: Patient safety is an important issue and receiving increasing attention. Information technology (IT) and IT-based strategies as the secondary use of already existing data within hospital information systems can help to improve patient safety. Objective: To investigate experts' knowledge and opinions regarding relevant patient safety problems, their measurability in existing clinical routine data and potential challenges in the field. We also wanted to get an overview of already deployed indicators for patient safety. Methods: Semi structured interviews with 20 experts from different healthcare domains were conducted and analyzed using a qualitative content analysis methodology. Results: The expert interviews offered a deeper insight into patient safety and quarried relevant patient safety problems including possibilities to measure them. The most often mentioned indicators were infection, complication and pressure ulcer. Conclusion: From an experts' perspective there are several challenges but equally a high potential for improving patient safety by the use of health IT.

Keywords. Patient Safety, Health Information Systems, Qualitative Research.

1. Introduction

Patient safety is an important global health issue and receiving increasing attention within different social spheres, e.g. governance, science and research [1-5]. Nevertheless, the amount of relevant patient safety matters is still estimated as high and associated with a burden for the patients and big expenditure for the affected health care systems [2,5].

In general, health IT can help to improve patient safety [6]. Especially, with regard to the rising quantity of information within the health care sector [7,8], big data and the secondary use of clinical routine data has gained intensifying interest [8] in this field. Compared with other approaches, e.g. voluntary reports or critical incidence reporting systems (CIRS), the reuse of existing clinical data within hospital information systems can help to overcome major disadvantages, e.g. a low sensitivity [9].

However, big health data research as well as patient safety research are both characterized by a high complexity [10] and it is often indeterminate what patient safety implies on an individual or a system level [1]. Several ways for measuring patient safety were proposed, and several indicators have been proposed, but best practice approaches or commonly accepted patient safety indicators are still missing [1].

Thus, the objective of this study was to investigate experts' knowledge and opinions regarding relevant patient safety problems, their measurability in existing clinical routine

[1] Corresponding Author: Michael Schaller, University for Health Sciences, Medical Informatics and Technology (UMIT), Eduard-Wallnöfer-Zentrum 1, A-6060 Hall in Tirol, E-Mail: michael.schaller@umit.at

data and potential challenges in the field. We also wanted to get an overview of already deployed indicators for patient safety.

2. Methods

Based on a broad literature analysis, a semi-structured interview guideline was developed and pretested to ensure its applicability for the expert interviews. The interviews were targeted on the following thematic blocks, always with regard to secondary use of clinical data:

- Experts' understanding of patient safety and relevant patient safety problems
- How to measure patient safety
- Suggested patient safety indicators and whether they can be derived from clinical routine documentation
- Challenges regarding patient safety

To find the experts we used a 'theoretical sampling' approach [11]. The sample consisted of German-speaking experts in the field of patient safety, quality and risk management as well as health care professionals from different domains and disciplines. We invited experts from different hierarchical levels to get a broader picture.

The interviews were transcribed verbatim and analyzed computer-assisted using a qualitative content analysis methodology according to Kuckartz [12] and MAXQDA12 (verbi GmbH) as software tool.

3. Results

Up to now 20 interviews were conducted. Our preliminary result set includes 15 fully transcribed interviews. The average interview duration was about 29 (±11) minutes with a range between 16 and 48 minutes. The expert panel comprised two nursing scientists (Austria, A), a member of the section nursing management in a university hospital (A), two nursing directors from hospitals (A) and a nursing informatics specialist (A). Also, three senior physicians from whom one is responsible for the risk management and the chief physician from a university hospital (A) participated. In addition, a junior physician (A), an employee from the quality management of a hospital (Switzerland, CH), a quality management leader (Germany, G), one leading person within the patient representation/advocacy (A) and a health department official (A) participated.

3.1. Experts' understanding of patient safety and relevant patient safety problems

When asked to give their definition or understanding of patient safety the experts often referred to established definitions, concepts, models or taxonomies. They named e.g. definitions from expert associations, as the German Medical Center for Quality in Medicine (Ärztliches Zentrum für Qualität in der Medizin, ÄZQ). A common baseline-definition which could be derived from the experts within almost all interviews was 'avoiding preventable harm, avoiding errors, preventing adverse events'. However, besides this common core definition, the experts reported a wide variety of personal definitions and understandings of patient safety depending upon their backgrounds, e.g. biological-psychological-social approach complemented by a spiritual dimension. In

addition a systemic patient safety perspective was described implying interactions within different sociotechnical or information systems and their components, stakeholders and processes.

Overall the construct 'patient safety' was described as highly relevant for the health care system and has to be tackled comprehensively and holistically. According to the experts' opinion patient safety had to include both, the quality dimensions (structure, process and outcome) as well as diverse other system components and their interplay.

When asked for relevant and concrete patient safety problems, the experts listed a variety of different problems that could be clustered according to the three quality dimensions.

- In the dimension of structure quality, problems or aspects like transport safety, staffing, training, education and experience were mentioned. Furthermore the patient safety attitude and awareness of health care professionals plus supporting staff were reported as well as the patient safety culture. Another structural problem from the experts' point of view was the poor consideration of the psychological perspective within the patient safety topic.
- In the dimension of process quality, patient safety problems affected social interaction (staff-to-patient, e.g. for building trust; staff-to-staff) and documentation (including trans-, inter- and multidisciplinary and inter-professional aspects). Explicitly, interface or communication problems as source of possible patient safety incidents were named, especially with regard to poor semantic interoperability of different information systems or impeded information flow. However, they pointed out that such problems would be hard to measure. Another relevant problem field named was medication safety with reference to interactions, side effects and similarity problems (look-alike, sound-alike), as well as prescription, transcription and dispensing errors.
- In the dimension of outcome quality, reported problems were visible in the core definition of patient safety: preventable harm and adverse events. Here, one commonly named problem was to quantify the outcomes, e.g. the ratio of pneumothoraxes, infections in order to benchmark the quality of a special domain. Furthermore, the need for ratiocination about such indices within the (interdisciplinary) health care team or by individuals was mentioned by several experts, i.e. drawing conclusions from the outcome dimension about the processes that led to the outcomes, i.e. the monitoring of the patients in long-time perspectives.

From the experts' point of view the interaction between the quality dimensions were named e.g. the recognition of context factors (e.g. structural aspects) which could influence e.g. fall rates or the disinfection-rate by measuring only the consumption of disinfectant.

3.2. How to measure patient safety

We received a heterogeneous response spectrum to the question how to measure patient safety. On the one hand there was no consent between the experts whether patient safety is measurable anyway. On the other hand a wide range of possibilities to gather relevant aspects of patient safety were named.

From the experts' point of view several structural aspects that affect patient safety were named for measuring: e.g. transport safety (e.g. defibrillator in the elevator – yes vs. no), staff dependent factors (e.g. quantity, experience, training, education), aspects

that affect the construction (e.g. barrier free – yes vs. no) and whether relevant diagnostic and therapeutic methods are provided within a hospital. From a nursing-perspective it was mentioned that context factors are systematically gathered, e.g. with regard to falls in the hospital, i.e. among others where and when the fall occurred and additionally there can be free-text boxes for amendments.

Often the experts proposed to disassemble processes in more little steps for measuring relevant aspects (e.g. standard operating procedures, checklists, guidelines, detection whether everything is accomplished). If these steps are documented electronically, process indicators could easily be derived (e.g. filling rate in checklists for a perioperative antibiotic prophylaxis). Furthermore, the time dimension could be used for deriving indicators as the timeline is represented within clinical data. Easily timestamps of admission, diagnoses, start of the state of the art therapies, interventions and durations between these time points or durations of whole sub-processes, e.g. the duration of a surgical operation or the duration of hospital stays, could be extracted from the data. With regard to medication safety from the experts' point of view it is necessary to have IT-based systems to control the prescriptions and to measure medication interactions or implications of specific medications to patient safety (e.g. different drugs with similar side effect profiles, such as prolongation of QT-times). Additionally, the quality of the documentation was named as a relevant problem that should be measured. The experts stated that it is hard to measure relevant patient safety aspects as attitude, communication and social interaction or even a patient safety culture. For the acquisition of the relevant processual data, information from audits, nursing rounds or case reviews were mentioned. Documentation quality could be assessed by the use of chart reviews. One expert suggested to measure number and impact of standards which are applied in a specific health care organization.

In addition to this, the experts named several aspects within the field of patient safety that could be measured from an outcome perspective. Here they mentioned a big variety of different parameters, e.g. the prevalence of harm: complications (infections, hemorrhage, neurological deficits, paralysis, disorders of brain capacity), pain and fear and irritations, falls (depending on context factors), decubital ulcera and suicides. Specific infections were e.g. catheter-associated infections, ventilator-associated pneumonias, and according to the experts these infections, other complications or incidents (e.g. falls) could be recorded and monitored systematically and standardized. Additionally, these rates and their context factors could be used for benchmarking processes. It is mentioned in the interviews that differences within the measuring process could lead to problems for the comparability. Other named relevant and measurable outcome-parameters are quality of life and patient satisfaction. But this implies that the respective parameters are collected and documented or even evaluated in regularly follow ups. Nursing experts often mentioned the assessment and screening of risk factors, e.g. with regard to fall and decubiti the Morse Fall and the Braden scale. The scores could then be compared to outcomes.

Furthermore, CIRS reports and the number of lawsuits could be used to capture concrete patient safety aspects. One expert added scientific output of an organization as patient safety parameter, i.e. a publication rate and the number of PhD-programs in a university hospital.

Table 1. Suggested patient safety indicators. The abbreviation 'n.a.' denotes 'not available', 'not applicable' or 'no answer'

Patient safety indicator	Type (structure, process, outcome)	Identifiability in clinical routine data
Fall/Fallrate (falls per thousand nursing days)	outcome	yes
Pressure ulcer	outcome	yes
Death in low mortality diagnosis related groups (DRG)	outcome	n.a.
Mortality	outcome	yes, partially long-time data capture
Lethality	outcome	yes, partially long-time data capture
Morbidity	outcome	yes, partially long-time data capture
Suicide	outcome	n.a.
Quality of life	outcome	n.a.
Patient satisfaction	outcome	n.a.
Injury index (injuries relative to falls)	outcome	yes
Fracture index (fractures relative to falls)	outcome	yes
Length of stay	outcome	yes
Complication rate (depending on discipline and complication)	outcome	yes
Infection rate (depending on discipline and infection)	outcome	yes
Ventilator-associated pneumonia	outcome	yes
Preventable adverse medication event	outcome	n.a.
Re-operation	outcome	yes
Unplanned inpatient re-admission within 30 days	outcome	yes
Delirium rate	outcome	yes
Pain	process/outcome	yes
Wrong site surgery	process/outcome	n.a.
Postoperative infection progression	process	n.a.
Perioperative antibiotics prophylaxis	process	n.a.
Germinal situation	process	n.a.
Activities for the restriction of freedom	process	yes
Number of staff away sick	structure	n.a.

3.3. Suggested patient safety indicators and whether they can be derived from clinical routine documentation

Table 1 summarizes the specific named concrete patient safety indicators, clusters according structure, process respective outcome dimensions and analysis whether the indicator can be identified within clinical routine documentation.

3.4. Challenges regarding patient safety

Several challenges are named by the experts, e.g. the merging and depiction of the information related to patient safety from diverse information systems into one information system or data platform by improving the semantic interoperability. This

would lead to an electronic health record which pools the information and supports the interface information flow and work flow, e.g. for the avoidance of (transmission) errors. A challenge is also medication safety that needs to be improved with the support of IT, e.g. through alerting systems for medication side effects and interactions. The development of suitable early warning or decision support systems is especially mentioned by the experts with a medical degree.

For the development of all IT-based instruments it was recommended to consider the experience and the assessment of the future users within the developing process.

One expert saw a challenge in the development of intelligent algorithms for filtering out artefacts for the secondary use of clinical routine data.

Data and information safety and the 'transparent patient' were also named as problems that have to be handled. Besides, the experts mentioned the challenge to use IT as an instrument without losing the humanity respectively the knowledge about e.g. diseases, therapies.

4. Discussion

The expert interviews offered a deeper insight into patient safety and quarried relevant patient safety problems including possibilities to measure them. The interviews generated a list of relevant patient safety indicators and an analysis of the possibility to align them with available clinical data (of the secondary use). The experts also named additional, but also important patient safety aspects which are difficult to extract from the already available clinical routine data, e.g. construction measures or the patient safety attitude of the staff.

In order to allow secondary use of clinical data to derive patient safety indicators, there are several conditions that have to be fulfilled, i.e. that the documentation has to be executed electronically, completely and structured. In addition the relevant data have to be accessible.

The indications of the experts concerning alternative ways to measure patient safety showed a heterogeneous spectrum. On the one hand they named case-based methodologies e.g. individual case reviews and nursing rounds, screenings, risk assessments and CIRS. However, these procedures are connected with a huge effort and a lack of evidence whether they are appropriate for improving patient safety [13]. On the other hand, the experts described specific systematic and standardized data collection over the time especially with regard to outcome parameters, e.g. infections and complications.

Overall the preliminary results point the complexity of the patient safety topic [10] out and they advert to the need for a clear definition of patient safety. The statements of the experts indicate that the construct patient safety has to be considered from different levels and taking into account the local context, before deciding on suitable patient safety indicators.

Some limitations of this study have to be pointed out. Only German-speaking experts were interviewed. We used purposeful sampling to identify the experts. Further expert-interviews are in preparation and additional experts will be recruited. This extension of the expert panel includes a quality and patient safety manager (CH), a quality and clinical risk manager (G), a senior scientist (G) and two members of relevant patient safety associations from Austria and Germany for completing the picture of

patient safety. The interview sessions will be pursued until data saturation is accomplished.

As a next step, the results of the expert panel will be triangulated with patient safety indicators which are extracted from literature. Then, a set of patient safety indicators will be selected, tested and validated based on available clinical routine data. By this, we will be able to show whether secondary use of clinical data is a promising approach to measure patient safety based on indicators.

5. Acknowledgements

We would like to acknowledge all participating experts and institutions, among others the Austrian Nursing Association (ÖGKV). This work takes place within the PATIS-Project (A PATient safety Intelligence System and framework for the secondary use of multimodal clinical data to assess and improve patient safety) which is supported by the Austrian Science Foundation (FWF – P 29076).

6. References

[1] M. Geraedts, S.E. Drösler, K. Döbler, M. Eberlein-Gonska, G. Heller, S. Kuske, T. Manser, B. Sens, J. Stausberg, M. Schrappe, DNVF-Memorandum III „Methoden für die Versorgungsforschung“, Teil 3: Methoden der Qualitäts- und Patientensicherheitsforschung, Gesundheitswesen 79 (2017), e95-e124.

[2] J.-M. Guise, M. Hansen, W. Lambert, K. O'Brian, The role of simulation in mixed-methods research: a framework and application to patient safety, BMC Health Services Research 17 (2017), 322. doi: 10.1186/s12913-017-2255-7

[3] J.J. Kadhim, Patient Safety in Primary Care: A Concept Analysis in Nursing, Kufa journal for nursing science 7(1) (2017), 17-28.

[4] M. Ock, S.Y. Lim, M.-W. Jo, S.-i. Lee, Frequency, Expected Effects, Obstacles, and Facilitators of Disclosure of Patient Safety Incidents: A Systematic Review, J Prev Med Public Health 50(2) (2017), 68-82.

[5] World Health Organization (WHO), Patient Safety: Making health care safer, World Health Organization, Geneva, 2017, http://apps.who.int/iris/bitstream/10665/255507/1/WHO-HIS-SDS-2017.11-eng.pdf?ua=1, last access: 29.1.2018.

[6] F. Magrabi, M.-S. Ong, E. Coiera, Health IT for Patient Safety and Improving the Safety of Health IT, in: Evidence-Based Health Informatics. IOS Press, Amsterdam, 2016. pp. 25-36. doi:10.3233/978-1-61499-635-4-25

[7] M. Rigby, E. Ammenwerth, The Need for Evidence in Health Informatics, in: Evidence-Based Health Informatics. IOS Press, Amsterdam, 2016. pp. 3-13. doi:10.3233/978-1-61499-635-4-25

[8] W.O. Hackl, E. Ammenwerth, SPIRIT: Systematic Planning of Intelligent Reuse of Integrated Clinical Routine Data. A Conceptual Best-practice Framework and Procedure Model, Methods Inf Med 55 (2016), 114-124.

[9] Sachverständigenrat zur Begutachtung der Entwicklung im Gesundheitswesen, Gutachten 2007. Kooperation und Verantwortung - Voraussetzungen einer zielorientierten Gesundheitsversorgung, http://dipbt.bundestag.de/dip21/btd/16/063/1606339.pdf, last access: 23.04.2017.

[10] S.P. Rodrigues, N.J. van Eck, L. Waltman, F.W. Jansen, Mapping patient safety: a large-scale literature review using bibliometric visualisation techniques, BMJ Open 4(3) (2014), e004468. doi:10.1136/bmjopen-2013-004468

[11] A. Bogner, B. Littig, W. Menz, Interviews mit Experten. Eine praxisorientierte Einführung, Springer, Wiesbaden, 2014.

[12] U. Kuckartz, Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung, Beltz Juventa, Weinheim, 2012.

[13] S. Kristensen, J. Mainz, P. Bartels, Selection of indicators for continous monitoring of patient safety: recommendations of the project 'safety improvement for patients in Europe', International Journal for Quality in Health Care 21(3) (2009), 169-175.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-307

HEALTHeBIKES – Smart E-Bike Prototype for Controlled Exercise in Telerehabilitation Programs

Patrick MAYR[a,b,1], Markus FALGENHAUER[a], Robert MODRE-OSPRIAN[a], Peter HOFMANN[c], Heimo TRANINGER[d], Michael RATH[e], Günter SCHREIER[a,b]
[a]*AIT Austrian Institute of Technology. Graz*
[b]*Institute of Neural Engineering, Graz University of Technology, Graz*
[c]*Exercise Physiology, Training and Training Therapy Research Group, Institute of Sports Sciences, University of Graz, Graz*
[d]*ZARG Zentrum Ambulante Reha Graz, Graz*
[e]*EBIKE EXPERTS EUROPE Limited, Gabersdorf*

Abstract. E-Bikes in telerehabilitation programs could be a new intervention for more sustainable rehabilitation results. The aim is to design and build a prototype of an E-Bike usable for rehabilitation – a HEALTHeBIKE. It should avoid over-exercising, work independently of the environment and it should enable cycling in a group despite different reference exercise intensities. To achieve these goals, requirements for this system architecture have been identified. A system architecture including an Arduino microcontroller, an Android smartphone and a telemonitoring platform was presented. A power output regulated proportional-integral controller to adjust the motor assistance has been implemented. A feasibility study with two subjects cycling in a group was performed. Seven test rides on varying terrain (flat, hilly, mountainous and uphill) with the same and different exercise intensities were completed. The mean power output was close to or below the target power output of the cyclist for all test rides with a maximal error of 6.7 % above and 27.6 % below the target. Although the exercise intensities of the two subjects were clearly different, cycling in a group was possible without over-exercising.

Keywords. E-Bike, electric bike, power output control, telerehabilitation

1. Introduction

Prescribing physical exercise as medicine is a key element in the rehabilitation of chronic diseases [1]. For efficient and safe physical exercises, it is important to appropriately choose the exercise intensity. In rehabilitation programs, stationary cycling ergometers are widely used for physical exercise. Individual power output based training profiles are applied to achieve the individual recommended exercise intensity [2,3]. However, most rehabilitation programs have in common that after initial success they are not further pursued [4]. Therefore, new interventions are necessary to secure sustainable effects.

An approach for sustainable rehabilitation could be based on eHealth and mHealth closed loop solutions, in which the health status of the patient is continuously monitored and the therapy is adapted as needed by healthcare professionals. Various groups have

[1] Corresponding Author: Patrick Mayr, AIT Austrian Institute of Technology, Reininghausstraße 13/1 8020 Graz, E-Mail: pat.mayr@outlook.com

demonstrated the effect of tele-training on long-term adherence [5,6]. In a previous work, we have developed a closed loop system consisting of a cycle ergometer, an Android Smartphone and a backend telehealth platform [7].

Electric bicycles (E-Bikes) are growing in popularity and are used for daily transportation and leisure activities. Compared to stationary cycling ergometers, E-Bikes can be used outside, in unrestricted groups and as part of daily life activities. Usual commercial E-Bikes have multiple assistance levels, which can be changed manually. On a stationary ergometer, it is easily possible to maintain and alter the exercise intensity by regulating the workload. Indicators for the exercise intensity can be the power output of the cyclist and the heart rate. However, on an E-Bike, more degrees of freedom (e.g. road gradient, gear transmission ratio, wind) must be considered. Continuous adaptions of speed and motor assistance are required to reach and maintain the desired exercise intensity. Such adaptions currently need to be done by the cyclist. Automatic regulation of the motor assistance could support the cyclist in maintaining the required exercise intensity and prevent over-exercising.

Meyer et al. [8] presented such a system which used a sliding mode controller and a feedforward controller to keep the heart rate constant and achieved good tracking performance. Corno et al. [9] developed a heart rate regulated bicycle which kept the heart rate within 10 bpm by adjusting the transmission ratio. However, with this approach cycling in a group with individual intensities is not possible. Furthermore, the heart rate shows a non-linear relation to the workload and it is dependent on several additional parameters, such as inter-subject variability or exogenous factors (temperature, humidity) [10,11]. Therefore, for such approaches, an individual heart rate model is required. The power output of the cyclist is a widely used parameter for the evaluation of the exercise intensity, especially for competitive cycling. The power output is instant whereas the heart rate responds with some delay. Therefore, a power regulated control system could lead to a more intuitive regulation and faster reactions, which results in a better riding experience.

Aim of this paper is to design and build a prototype of an E-Bike usable for rehabilitation – a HEALTHeBIKE which fulfils the following requirements: (a) avoid overload by keeping the power output of the cyclist at or under a desired intensity level, (b) work independently of the environment and (c) allow cycling in a group with different individual exercise intensities.

2. Materials and Methods

2.1. Prototype components

The commercial E-Bikes crossroad inframe and wellness inframe (EBIKE EXPERTS EUROPE Limited, Austria) were used and modified for this study. The E-Bikes were equipped with a 500 Watt rear hub motor connected to a motor controller, a speed sensor and a torque sensor. The motor controller featured a throttle interface, which required a voltage signal as input. This throttle interface was used to adjust the motor assistance.

The power output of the cyclist and the heart rate were chosen as indicators for the exercise intensity. With Polar H7 heart rate chest straps (Polar Electro, Finland) heart rate measurements were performed. Built-in bottom bracket torque sensors were used to measure torque applied by the cycler and angular velocity of the pedal crank. The product

of torque * angular velocity was used to calculate the power output applied to the pedals by the cyclist.

An Arduino microcontroller was used to capture and process sensor signals from the speed and bottom bracket torque sensor. By that, the speed of the bike and the cadence and power output of the cyclist could be obtained during riding. Furthermore, the Arduino microcontroller was used to set the motor assistance through the throttle input of the E-Bikes motor controller. The Arduino microcontroller provided a pulse-width modulated voltage signal (PWM) which was low pass filtered to get the desired voltage level. Through this approach, a continuous motor assistance was achieved.

An Android smartphone hosted the control system and acted as display. Furthermore, it was used for collecting data and communicating with the telemonitoring platform. The E-Bike was integrated into a closed loop rehabilitation scenario, where the training sessions could be monitored and adapted as needed.

To calculate the power output of the cyclist with the signal provided by the built-in torque sensor additional steps were necessary. The torque signal was sampled in 18 degrees steps of the pedal position, resulting in 20 measurements per cycle. For a physically meaningful interpretation, the torque signal was calibrated with a Garmin Vector power meter (Schaffhausen, Switzerland). Simultaneous measurements of the build-in torque sensor and a Garmin Vector power meter while increasing the power output from 30 W to 300W were performed and a correlation coefficient of 0,93 was found. Through least-mean-square-fitting, a transformation from the voltage signal of the torque sensor to a representing torque signal was then obtained. In case, the cyclist did not perform a full rotation within 2 seconds, the output power was set to 0 W.

To comply with the Austrian jurisdiction, the motor assistance had to be stopped if the 25 km/h limit was exceeded or the cyclist stopped pedalling.

2.2. Feasibility study

Feasibility of controlled exercise in a group on the prototype was evaluated with two healthy subjects cycling together on the two modified E-Bikes. Subject 2 was defined as the “leader” and subject 1 as the “follower”. Seven test rides were performed which included flat, hilly, mountainous and uphill terrain. The subjects performed two test rides with the same target power and six test rides where the target power was different by minimal 40 Watts and maximal 60 Watts. The target power was varied for the subjects and test rides from 60 Watts to 220 Watts. The subjects were permitted to cycle at a cadence of 60-70 rpm. In order to simulate an unexperienced cycling behaviour, shifting the transmission ratio had to be changed as little as possible. Subject 2 (“leader”) was requested to stay in maximal distance of 10 m behind subject 1 (“follower”).

3. Results

3.1. Prototype

The results of the prototype are split in two parts. First the system architecture is described, followed by the used control system.

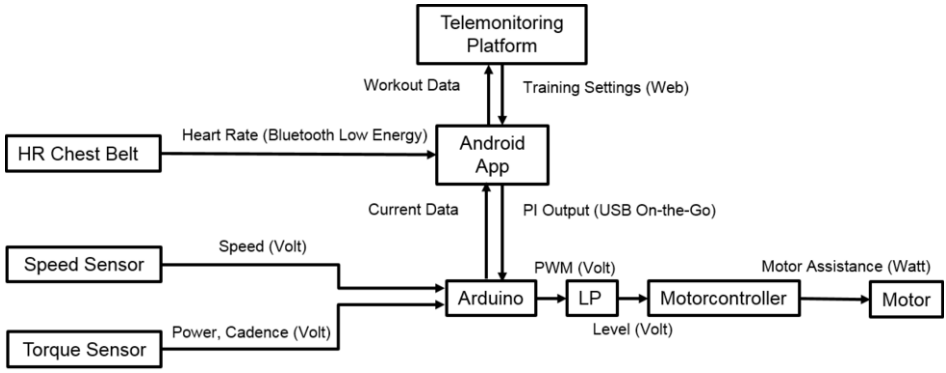


Figure 1. System architecture of the HEALTHeBIKES prototype

3.1.1. System architecture

The system architecture of the designed and built prototype can be seen in Figure 1. The Android smartphone was connected to the Arduino microcontroller via USB-On-the-Go (USB-OTG), to the heart rate chest strap via Bluetooth Low Energy (BLE) and to the AIT telehealth platform via web. The workout data and training settings were transmitted between the telemonitoring platform and the Android smartphone. Current speed, cadence and power data were captured by the Arduino microcontroller and sent to the Android smartphone. The control system hosted on the Android smartphone calculated the required motor assistance based on the current data and the training settings and sent to the Arduino microcontroller.

3.1.2. Control system

The cyclist was assumed to maintain the target velocity (via cadence and gear transmission ratio) through power output adjustments. The target velocity was defined as the personally preferred velocity or – when cycling in a group – the velocity of the group leader.

The controller was designed as a proportional-integral controller (PI controller), see Eq. 1. The output *u(t)* was a voltage level that was applied on the throttle input of the E-Bikes motor controller to set the corresponding motor assistance.

$$u(t_k) = u_0 + K_p \times e(t_k) + K_i \times \sum_{j=1}^{k} e_j(t_k) \times (t_j - t_{j-1}) \quad (1)$$

The error term *e(t)* was the difference between the desired reference power output of the cyclist and the current power output of the cyclist. The term u_0 was a default offset. Motor assistance started when then power output was higher than this offset. With the coefficients K_p and K_i the proportional and integral terms were weighted. A new power value was calculated after a full rotation of the pedal. Therefore, the sampling time $(t_j - t_{j-1})$ was the inversed cadence. Furthermore, an anti-windup method was implemented.

The coefficients $K_p$and K_i were determined through experiments with the goal of a fast settling time but without overshoot for good riding experience. K_p was set to 0.001 V/W and K_i to 0.0025 V/Ws.

Table 1. Evaluation of the performed test rides in a group of two as achieved during 7 test rides. Reference power output (Power ref), mean power output (Power mean) and the difference (Diff.) between these two values are shown.

			Subject 1			Subject 2		
No.	**Gradient**	**Duration**	**Power ref**	**Power mean**	**Diff.**	**Power ref**	**Power mean**	**Diff.**
#	-	min	Watt	Watt	%	Watt	Watt	%
1	flat	5.5	100	106.7	6.7	100	100.6	0.6
2	hilly	9.6	60	59.9	0.2	120	92.9	22.6
3	hilly	10.4	120	96.1	19.9	60	55.9	6.8
4	hilly	11.3	100	86.3	13.7	100	83.1	16.9
5	mountainous	23.6	80	68.1	14.9	120	107.9	10.1
6	hilly	15.4	100	96.6	3.4	140	125.1	10.6
7	uphill	6.7	160	161.6	1.0	220	221.3	0.6

Additionally, a heart rate limit was implemented. When the subject reached this limit, he/she was warned through the display and requested to reduce the workload or to stop exercising.

3.2. Feasibility study

In Table 1, the comparison of the desired reference power output (Power ref) to the mean power output (Power mean) for the two subjects as well as the difference between these two values for each test ride can be seen. In total, the difference between Power ref and Power mean ranged between 0.2 % and 22.6 %.

Subject 2 confirmed that it was possible to follow the group leader subject 1. The maximum distance to the group leader of 10 m during all the test rides was not exceeded.

The system behaviour is illustrated in Figure 2. Data from test ride 7 of subject 2 are shown. In the beginning, the power output of subject 1 was lower than the target of 220 Watts, since the road gradient was too low to reach this target at the group leader's target velocity. As the road gradient increased, the power output of the cyclist increased to a maximum of 331 Watt (50 % over the target power output). As soon as the power output

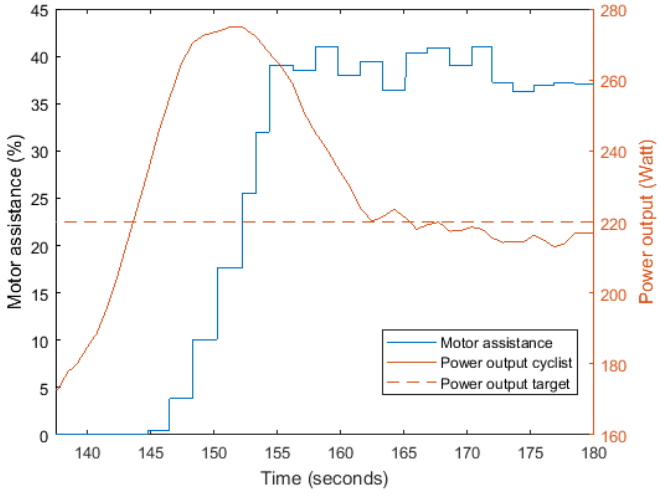


Figure 2. Regulation of the motor assistance due to changed power output of the cyclist of test ride No. 7 for subject 2. A moving average window of 10 s is applied to the power output of the cyclist.

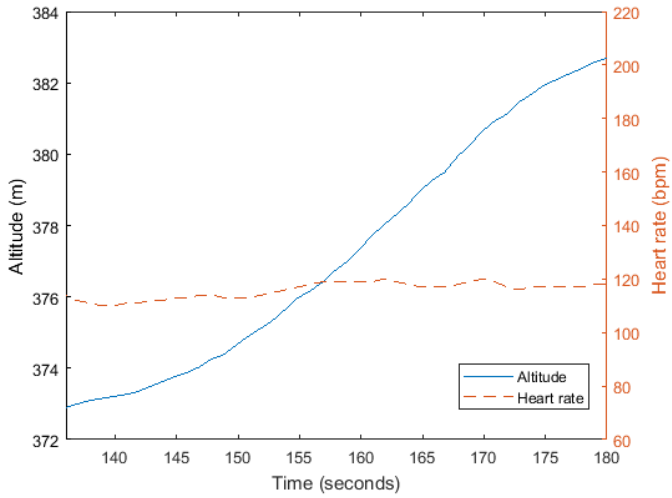


Figure 3. Altitude profile and heart rate of test ride No. 7 for subject 2

of the cyclist was above the target power, the motor assistance increased which can be seen at second 145. The motor assistance settled after around 15 seconds when the desired power output was reached. The heart rate of the subject stayed constant over this time indicating that the regulation was fast enough to avoid overloading of the subject.

4. Discussion

In this paper, a smart E-Bike for controlled exercising in telerehabilitation programs has been presented. An Arduino microcontroller and an Android smartphone were added to an E-Bike to gather and process the sensor data, to hosts a control system and to communicate with a telemonitoring platform. A PI controller regulated the motor assistance based on the difference between the reference power output of the cyclist and the actual power output.

Results show that the mean power output of the cyclists for the different test rides were mostly close or below the target power outputs and single peaks heavily above but short enough to be save. Therefore, based on mean power output, over-exercising was successfully prevented.

For flat and uphill tracks, the target power could be reached very closely. On hilly and mountainous tracks, however, the mean power output of the cyclists was lower than the necessary power output, especially for the subject with the higher reference power output. This can be explained by the downhill road sections, where the cyclists' power was lower than the target power. For the subject with the higher reference power it was earlier not necessary to provide the reference power while maintaining the cadence and gear selection. Additionally, restrictions because of traffic, road curves and intersections reduced the mean power. If a system to maintain the target power output even during downhill sections was required, an active braking element is required, e.g. a motor with recuperation could be used. However, we expect that – due to the resulting unusual cycling behaviour – this might have negative effects on the compliance.

Since the physiological responses are not only dependent on the mean power output of the cyclist, further analyses are required, e.g. if the system reacts fast enough to limit the amplitude and duration of power output peaks of the cyclist to secure a safe and efficient physical exercise.

In our example with strong increases in power output of the cyclist, we found a constant heart rate during this time, indicating that the system reacted fast enough in this situation for the subject to secure a safe cardio-respiratory strain below given limits. By changing the PI coefficients, the system dynamics can individually be adapted to each subject.

Since subject 2 could keep the distance to subject 1 within a range of 10 m in all test rides, we conclude that cycling in a group of two is possible for different intensity values, within the chosen intensity range and terrain. Cycling in a group may increase patient motivation and therefore increase the compliance.

As a next step, further studies with healthy male and female subjects as well as eligible patients from Phase III / IV cardiac rehabilitation of a local rehabilitation centre will be performed to validate and further improve the HEALTHeBIKES prototype in a cardiac rehabilitation scenario.

Acknowledgement

This work was partly funded by “Zukunftsfonds Steiermark” (GZ: ABT08-183057/2016 PN:8005).

References

[1] B. K. Pedersen and B. Saltin, “Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases,” Scand. J. Med. Sci. Sports, vol. 25, pp. 1–72, Dec. 2015.

[2] S. C. Smith et al., “AHA/ACCF Secondary Prevention and Risk Reduction Therapy for Patients with Coronary and other Atherosclerotic Vascular Disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation.,” Circulation, vol. 124, no. 22, pp. 2458–73, Nov. 2011.

[3] J. Niebauer, K. Mayr, M. Tschentscher, R. Pokan, and W. Benzer, “Outpatient cardiac rehabilitation: the Austrian model,” Eur. J. Prev. Cardiol., vol. 20, no. 3, pp. 468–479, Jun. 2013.

[4] Laimer H, “Ein Jahr nach stationärer Rehabilitation - was blieb?,” J. für Kardiol. -Austrian J. Cardiol., vol. 11, no. 11, pp. 437–440, 2004.

[5] K. Wallner et al., “Telerehabilitation in coronary artery disease (TRIC-Study), 12 months data,” Eur. J. Prev. Cardiol., vol. 23, no. 2, p. 51, 2016.

[6] D. Kairy, P. Lehoux, C. Vincent, and M. Visintin, “A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation,” Disabil. Rehabil., vol. 31, no. 6, pp. 427–447, Jan. 2009.

[7] W. Berthold, “Erweiterung einer Telehealth-Plattform für die Rehabilitation mit Ergometertraining,” Graz University of Technology, 2016.

[8] D. Meyer, W. Zhang, and M. Tomizuka, “Sliding Mode Control for Heart Rate Regulation of Electric Bicycle Riders,” in Proceedings of the ASME 2015 Dynamic Systems and Control Conference, 2015, vol. 2, p. V002T27A003.

[9] M. Corno, P. Giani, M. Tanelli, and S. M. Savaresi, “Human-in-the-Loop Bicycle Control via Active Heart Rate Regulation,” IEEE Trans. Control Syst. Technol., vol. 23, no. 3, pp. 1029–1040, May 2015.

[10] P. Hofmann, R. Pokan, S. P. von Duvillard, F. J. Seibert, R. Zweiker, and P. Schmid, “Heart rate performance curve during incremental cycle ergometer exercise in healthy young male subjects.,” Med. Sci. Sports Exerc., vol. 29, no. 6, pp. 762–8, Jun. 1997.

[11] P. Hofmann et al., “%HRmax target heart rate is dependent on heart rate performance curve deflection.,” Med. Sci. Sports Exerc., vol. 33, no. 10, pp. 1726–31, Oct. 2001.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-314

Two-Stage Evaluation of a Telehealth Nutrition Management Service in Support of Diabesity Therapy

Veronika SCHUSTERBAUER[a,b,1], Dieter FEITEK[c,2], Peter KASTNER[a] and Hermann TOPLAK[d]
[a]*AIT Austrian Institute of Technology GmbH, Graz, Austria*
[b]*Institute of Neural Engineering, Graz University of Technology, Graz, Austria*
[c]*Versicherung für Eisenbahnen und Bergbau (VAEB), Vienna, Austria*
[d]*Department of Internal Medicine, Medical University of Graz, Graz, Austria*

Abstract. Background: Diabesity, a combination of diabetes mellitus type 2 and obesity, is one of the biggest global health problems. Individual nutrition therapy and physical activity are effective measures in prevention and treatment. Methods: Requirements for an integrated diabesity service were specified and evaluated in two stages. The aim of the first stage was to perform a feasibility trial in patients with diabesity, using separate diaries. DiabMemory for diabetes and Nutrinaut for nutrition. Based on the results of the first stage, a prototype of an integrated diabesity solution (KIT-Nutriton, AIT) was developed and evaluated. Results: First stage trial was performed with 10 diabesity patients (2f, med:53 y. (IQR:7), T2DM, BMI > 28 kg/m²) with 3 months follow-up and a significant reduction of BMI. Participants noticed concerns about using two separate diary solutions. Results of the field trial with KIT-Nutrition with 14 healthy subjects (3f, med:26 y (IQR: 20,25)) showed that overall, 77.6 % of the intended tasks had been achieved. Conclusion: Results show that the integrated KIT-Nutrition app, providing access to a nutrition database, is feasible and accepted by the users. Before further trials can be made, an extension for regional food terms is recommended.

Keywords. Nutrition Assessment, Telemedicine, Obesity, Diabetes Mellitus, Android

1. Introduction

According to the *Global Burden of Disease Study 2015*, high body mass index (BMI) as well as high fasting plasma glucose are amongst the five leading global health risks [1]. Furthermore, the prevalence of obesity and diabetes mellitus type 2 (DMT2), diabesity in short, is rising exponentially all over the world. In 2013, there were about 2 billion overweight or obese people worldwide, of which about 20 – 25 % are expected to develop diabetes [2,3]. Both obesity and physical inactivity are established risk factors for developing diabetes [4,5.

There is also a pathophysiological connection between obesity and diabetes type 2 expressed in insulin deficiency and resistance. Apparently elevated levels of free fatty

[1] Corresponding Author: Veronika Schusterbauer, AIT Austrian Institute of Technology, Reininghausstraße 13/1 8020 Graz, E-Mail: v.schusterbauer@gmx.at
[2] Dieter Feitek and Veronika Schusterbauer contributed equally to this work with their Master's Thesis at the FH Joanneum Graz and the Technical University Graz, respectively.

acids in the blood plasma are one of the decisive conditions for development of an impaired glucose tolerance [6].

One of the most effective measures in prevention and treatment of diabesity is a lifestyle intervention, including medical nutrition therapy as well as physical activity [4]. Mobile applications are expected to have the power to bring about such a change in behavior, but there is a lack of culturally adapted and scientifically evaluated applications. A further demand, that cannot be met by usual nutrition applications, is offering a personalized diet for every individual [7-9]. Traditional methods in nutrition management, on the other hand, are usually biased towards a lower caloric intake as patients tend to underestimate the portion sizes of their meals [10].

In general, telehealth services are cost-effective and reliable methods in the treatment of chronic diseases [11]. Keep in Touch (KIT)[1] telehealth solutions, developed by the Austrian Institute of Technology (AIT), is a system that follows the closed loop healthcare principle. Its main building blocks are a health data server and a patient terminal, the latter comprising of a smartphone and a dedicated Android app building the front end. A web dashboard service for health professionals presents the back end [12].

One of the Disease Management Programs (DMP) working with the KIT system is the Health Dialogue Diabetes mellitus (HD-DM)[2] run by the Austrian Health insurance for Railway and Mining (VAEB). Patients from the HD-DM can voluntarily sign up to be part of DiabMemory[3], a telemonitoring program based on the KIT system. When patients get assigned to DiabMemory, they are provided with the needed measurement devices and a special training regarding their disease and usage of the equipment. Afterwards they are able to collect their vital parameters, like blood pressure, blood glucose and weight and activity, themselves. The healthcare professional in charge can then access the data via a web application and send feedback messages directly to the patients' phone [13].

The aim of our research is the inclusion of image assisted nutrition management, as a telehealth service, into the treatment of patients suffering from diabesity, alongside their usual DMP like HD-DM. This includes the design of a new treatment pathway as well as the adaption of the used telehealth system.

2. Methods

2.1. Part 1: Evaluation of the Acceptance of Electronic Nutrition Diaries

We planned a prospective study including ten patients suffering from diabetes mellitus type 2, that were part of the HD-DM program and the DiabMemory telemonitoring program. Inclusion criteria were, an age between 30 and 70 years and a BMI above 28 kg/m^2. Furthermore, it was obligatory that the patients took part in an inpatient treatment at a special treatment facility. Exclusion criteria were a dependency of the patient on insulin or a cardiac pacemaker. The study was approved by the ethics committee of the Federal State Lower Austria (No. GS4-EK-4/192-2012).

[1] AIT, KIT telehealth solutions: *https://kit.ait.ac.at/home/*, last access: 03.01.2018

[2] VAEB, HD-DM: *https://www.vaeb.at/portal27/vaebportal/content?contentid=10007.747216*, last access: 04.01.2018

[3] AIT, DiabMemory, *https://www.ait.ac.at/themen/telemedical-solutions/projects/diabmemory/*, last access: 04.01.2018

The telemedical, dietary intervention was conducted by using the two systems, DiabMemory (registered trademark by AIT Austrian Institute of Technology GmbH), and Nutrinaut (registered trademark by Dr. Dietmar Dörrer, Vienna, Austria), in combination. Nutrinaut is a web based system for image assisted dietary records, allowing the patient as well as the advisor to access the nutrition diary.

The follow-up period lasted for three months, starting after the stationary treatment. The therapy plan included the weekly recording of blood pressure, weight and blood sugar (3 values on one day). For the first month, the transmission of images of three main meals daily was obligatory. The patients received a weekly individual feedback from their dietitian as well as general motivational messages. Additionally, the body measures BMI and body fat as well as the HbA_{1c} were taken in the beginning and in the end of the intervention. For statistical testing, we used the Wilcoxon rank-sum test. In the end of the case series, the participants were interviewed regarding the benefit of the nutrition counselling [14].

2.2. Part 2: Feasibility Study for Inclusion of Nutrition into the KIT System

2.2.1. System

The requirements for the back end and front end of the nutrition management system, were determined in discussions with nutrition experts. A first version of the front end was implemented by extending the DiabMemory app using Android Studio (Google Inc., Mountain View, USA) and distributed for a closed user group under the name KIT-Nutrition. Its main functions, as depicted in Figure 1, are the ability to keep an electronic nutrition diary by adding meals including an image and the consisting food items, and the calculation of the basal energy expenditure using the Harris Benedict formula [15]. The energy expenditure during an entered activity was estimated based on a formula using the users' bodyweight and the metabolic equivalent of each activity [16,17].

For allowing the annotation of the meal images, a nutrition web service had to be set up which was written in NODE.js (NODE.js Foundation, San Francisco, USA) and published using Azure Cloud Services (Microsoft Azure, Microsoft, Redford, USA). The underlying database was the German *Bundeslebensmittelschlüssel*[1], which is the biggest scientific food composition database in Europe including about 15.000 entries, divided into 132 nutrients [18].

2.2.2. Field Trial

In order to test the application for usability and time consumption, we conducted a field trial with healthy volunteers consisting of work and university colleagues. Participants were invited to sign up for the beta version of the app on the Google Play Store[2] and were provided with a pseudonymized account for the KIT system.

The goal for every test user was to track and transmit his/her food intake for seven days, by taking a picture and adding the annotation of every meal. After a trial period of two weeks, the participants were asked to answer a questionnaire that was divided into

[1] Max Rubner Institut, Bundesforschungsinstitut für Ernährung und Lebensmittel, Karlsruhe, Bundeslebensmittelschlüssel V3.02, *https://*www.mri.bund.de, last access: 26.01.2018

[2] Google Inc., Play Store, *https://play.google.com/store*, last access: 03.01.2018

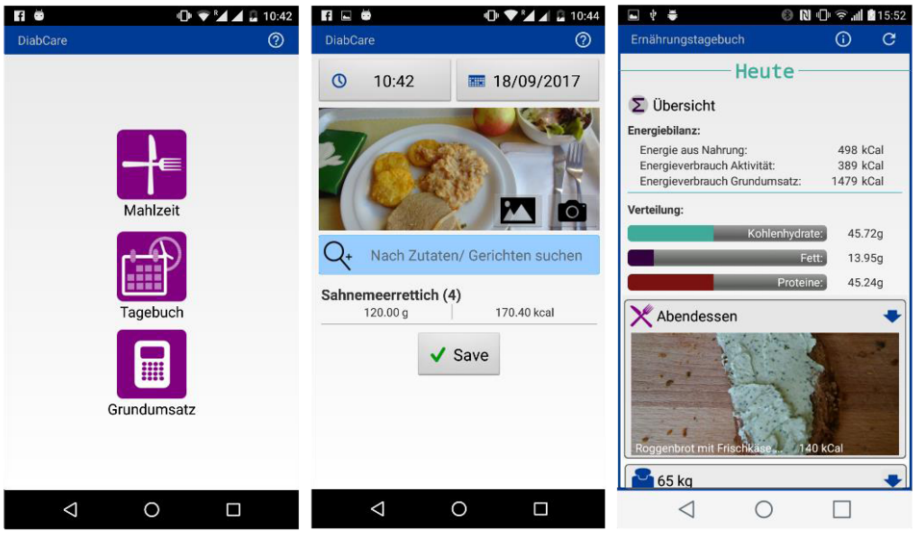


Figure 1: Overview of the KIT-Nutrition app. **Left:** Main activity **Middle**: Adding a new meal entry **Right:** The meal diary, showing the daily summary of calorie intake and usage.

two parts. Part one consisted of fourteen multiple choice questions graded according to the four-point Likert scale.

The questions were, for evaluation, separated into the sectors system quality (4 questions), information quality (3 questions), user satisfaction (4 questions) and net benefit (3 questions) guided by the *DeLone and McLean* model for assessing information system success [19,20]. For the different areas, the mean of the given answers was calculated. The second part of the questionnaire included open ended questions about the time consumption and the used phone, as well as the installed Android version and moreover, a section for problems and suggestions. The usage was calculated as the percentage of intended days of usage [21].

3. Results

3.1. Part 1: Evaluation of acceptance for DiabMemory and Nutrinaut

10 patients including two females, median age 53 years participated in the study. The median BMI at the beginning of the study, was 34,45 kg/m^2. Further characteristics of the study group are summarized in Table 1. It also shows, that the median BMI was reduced by 2,74 kg/m^2, which means a significant reduction ($p \leq 0{,}005$).

In total, the test subjects transmitted 79,6 % of the predefined 90 images per person, the median was 100% (min: 26%, max 100%). In total, the patients took images of 915 meals. There were 22 varying feedback messages sent by the dietitian.

The evaluation of the acceptance showed that 10 out of 10 participants thought, that the inclusion of nutrition counselling into the HD-DM is a good idea and that they would recommend the use of such a system to a person having the same kind of diabetes. 8 patients even stated that they would like to participate in the nutrition counselling for a longer period of time and all of the participants at least rather agreed that keeping a nutrition diary by taking pictures was easy (Figure 2).

Table 1: Overview of the basic data of the patients, at the beginning and the end of the study.

	Beginning	End	Significance
Patients: number (m/f)	10 (8/2)	10 (8/2)	-
Age in years: median (IQR)	53 (7)		-
BMI in kg/m²: median (IQR)	34,45 (6,18)	31,71 (5,18)	$p \leq 0{,}005$
HbA_{1c} in %: median (IQR)	5,9 (1,15)	6,05 (0,88)	n.s.
Body fat in %: median (IQR)	32 (6,25)	29,95 (6,63)	n.s.

In the end of the study, the only suggestion for improvement that was mentioned by several patients (6 out of 10) was the inclusion of the nutrition diary into the DiabMemory app, because of the remarkable effort needed for parallel management of two systems.

3.2. Part 2: Feasibility study

15 people signed up for the field trial. 1 person was excluded, because no Android phone was available, leaving 14 people including 3 females, with a median age of 26 (IQR: 20,25). All the participants used their smart phones on a regular basis.

In total, there were 407 images of meals and 1 181 annotations transmitted to the health data server. Out of the 14 volunteers, 9 reached the goal of using the app for seven days. In total, the usage resulted to 77,6 % of the intended goal. The return rate of the questionnaires, was 85,7 % (12 out of 14), reporting the use of 11 different phones. It took the test users about 5 (IQR: 2,25) minutes to enter one meal and about 18 (IQR: 14,83) minutes to enter the meals of the whole day (med (IQR)).

As summarized in Figure 3, left, the scores reached were 2,9 for system quality, 3,4 for information quality, 3,1 for user satisfaction and 2,9 for net benefit. The most frequently reported problems and suggestions (Figure 3, right) were problems with showing the meal diary and problems with finding the correct food items (9 out of 12). 7 participants uttered the wish to allow editing and deleting of entered measurements. The suggestions to use a more regional database and to allow for more personalization, e.g. better suggestions based on logged meals or the ability to save favorites, were both mentioned 6 times. 4 test users experienced app crashes and temporary problems with server connectivity.

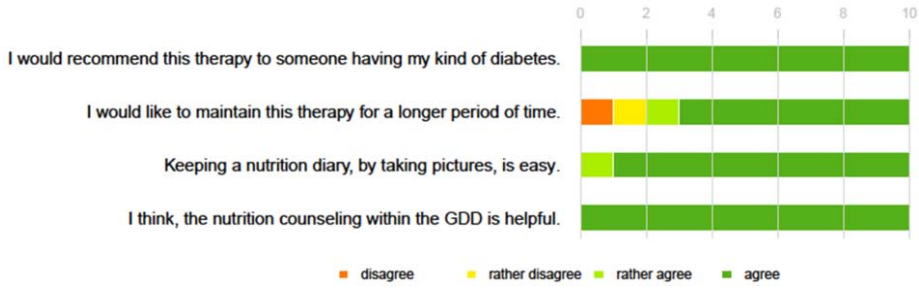


Figure 2: Results of chosen questions, about the acceptance of a nutritional counseling

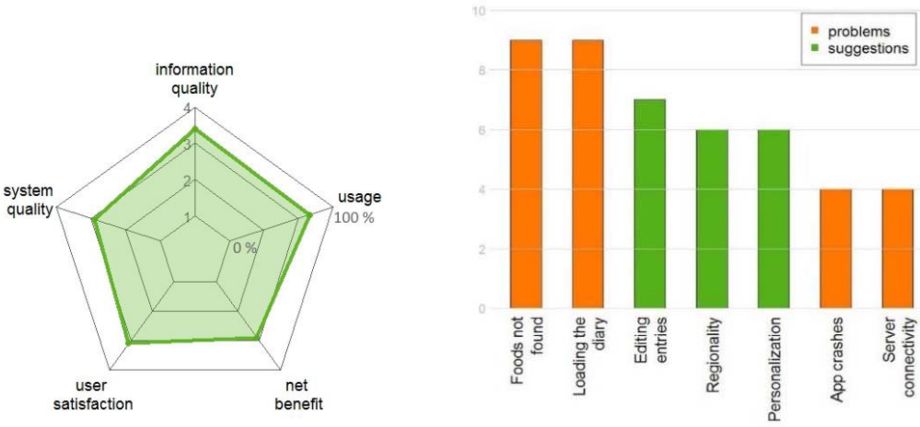


Figure 3: Left: System Success: information quality, system quality, user satisfaction and net benefit rated on a scale from 1 to 4, usage rate on a scale from 0 to 100 %**. Right:** Most common problems (orange) and suggestions (green).

4. Discussion

The results of the field trial were mainly positive, showing that the technical implementation of the system is feasible. The integration of the new type of diary data into the KIT system was conducted and so far, no problems were encountered. The success rate of more than 75% shows that the utilization of the diary app for a whole week, in a real-life setting, is possible. Based on the discussion with dietitians, a nutrition diary including about 3 weekdays and the weekend is already helpful for the application to nutrition counseling. The app worked stably on the majority of used phones although some difficulties were reported.

As stated, the field trial of the new system was conducted with the help of work and study colleagues, therefore the results of the questionnaire have to be treated with a certain amount of reserve, but the earlier evaluation study showed that the motivation for keeping an electronic nutrition diary among people affected by diabesity is very high. They also felt a high personal benefit from using the system. Therefore, a similar outcome is hoped for in case of a deployment of KIT-Nutrition in an applicable DMP.

The technical problems, which were mentioned by the test users, are of different origin. The problems with synchronization to the backend were due to partial unavailability of the database in use by the staging server. They were fixed by manually synchronizing the affected measurements. However, the client-server model based system architecture prevents data loss and guarantees data integrity even during temporary unavailability of the backend server. The other bugs concerned the Android app itself. For one of the phones used, the meal diary part of the app broke down. Why this happened exactly is unclear, but the Android version in use (Android 4.1) was beneath the target version (Android 5.0) of the KIT-Nutrition app. The other technical issues reported, present very valuable information to us and will be solved before the next app release.

The BLS was chosen as a food composition database because it is offered free of charge for scientific projects inside Europe, but apparently it does not quite match the needs of a common user in Austria. From the given feedback, it can be concluded that a

more comprehensive and regional database would be preferable, as people had troubles finding matching food items. One option would be an inclusion of the *Österreichische Nährwerttabelle (ÖNWT),* the Austrian extension of the BLS. The ÖNWT would enhance the usability as it includes Austrian synonyms of food items and meals and brand names of commercially available foods, without compromising the quality of information. For reasons of data integrity, it was impossible for patients to alter or delete the measurements they entered. This was decided upon the advice of healthcare professionals. According to the outcome of the field trial, this decision should be reconsidered as many test users were unhappy with the fact that they could not edit the meals they had entered.

According to nutrition experts working in the field of diabetes the possibilities of using such a system are numerous, the most important ones being:

- Treatment and prevention of obesity, diabetes, diabesity, or cardiovascular diseases by providing remote support for lifestyle intervention based on enhanced diary data including nutrition
- Virtual support for teaching programs for primary, secondary or tertiary prevention with focus on lifestyle, nutrition and physical activity.
- Inference of useful information for estimating the dosage in diabetes patients depending on insulin medication

Using image assisted dietary assessment would give dieticians the possibility to evaluate the patients' self-assessment and the integration into a telehealth system would further allow them to give prompt feedback, which is assumed to enhances the adherence of patients as it keeps their motivation up.

4.1. Outlook

In order to further verify the use of the system and to learn about possible long-term effects a randomized trial with people suffering from diabesity has to be conducted. As patients with diabesity are typically older and less technically experienced as our last group of test users they might face problems that we have not thought about yet. Before such a study can be started, the system must be further finalized. Especially the needed features for the backend have to be discussed and implemented.

Acknowledgement

We express our thanks to the Max Rubner Institut in Karlsruhe, Germany for providing us with the Bundeslebensmittelschlüssel Database for this research project.

References

[1] GBD 2015 Risk Factors Collaborators, "Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015," Lancet, vol. 388, pp. 1659–1724, 2016.

[2] Y. M. K. Farag and M. R. Gaballa, "Diabesity: an overview of a rising epidemic," Nephrol. Dial. Transplant., vol. 26, no. 1, pp. 28–35, Jan. 2011.

[3] S. Cuschieri, J. Vassallo et al., "The diabesity health economic crisis-the size of the crisis in a European island state following a cross-sectional study.," Arch. public Heal., vol. 74, p. 52, 2016.

[4] P. W. Sullivan, E. H. Morrato, and V. Ghushchyan, "Obesity , Inactivity , and the Prevalence of Diabetes andDiabetes-Related Cardiovascular Comorbidities in the U.S., 2000–2002," Diabetes Care, vol. 28, no. 7, pp. 1599–1603, 2005.
[5] D. R. Leitner, G. Frühbeck et al., "Obesity and type 2 diabetes: Two diseases with a need for combined treatment strategies - EASO can lead the way," Obes. Facts, vol. 10, no. 5, pp. 483–492, 2017.
[6] S. E. Kahn, R. L. Hull, and K. M. Utzschneider, "Insulin resistance and obesity," Nature, vol. 444, no. 14, pp. 840–846, 2006.
[7] Y. M. Khazrai, G. Defeudis, and P. Pozzilli, "Effect of diet on type 2 diabetes mellitus: a review," Diabetes. Metab. Res. Rev., vol. 32, no. 30, pp. 24–33, 2014.
[8] J. I. Recio-Rodríguez, C. Martín-Cantera et al., "Effectiveness of a smartphone application for improving healthy lifestyles, a randomized clinical trial (EVIDENT II): study protocol," BMC Public Health, vol. 14, pp. 1–13, 2014.
[9] S. S. Coughlin, M. Whitehead et al., "Smartphone Applications for Promoting Healthy Diet and Nutrition: A Literature Review.," Jacobs J. food Nutr., vol. 2, no. 3, p. 21, 2015.
[10] L. Gemming, J. Utter, and C. Ni Mhurchu, "Image-Assisted Dietary Assessment: A Systematic Review of the Evidence," J. Acad. Nutr. Diet., vol. 115, pp. 64–77, 2015.
[11] A. Gordon and J. Marchibroda, "Transforming Health Care : The Role of Health IT," 2012.
[12] AIT Austrian Institute of Technology GmbH, "Keep In Touch," 2017. [Online]. Available: https://kit.ait.ac.at/home/. [Accessed: 01-Oct-2017].
[13] I. Peinado, E. Villalba et al., "Strategic Intelligence Monitor on Personal Health Systems Phase 3 (SIMPHS 3)," 2015.
[14] D. Feitek, "Auswirkungen einer diätologischen Intervention mit telemedizinischer Nachbetreuung auf das Körpergewicht, die Körperfettmasse und den HbA1c von Patientinnen und Patienten mit Typ-2 Diabetes mellitus ohne Insulintherapie," FH Joanneum, 2013.
[15] D. C. Frankenfield, E. R. Muth, and W. A. Rowe, "The Harris-Benedict Studies of Human Basal Metabolism," J. Am. Diet. Assoc., vol. 98, no. 4, pp. 439–445, Apr. 1998.
[16] J. Hannon, L. Ex, and A. Gtm, "2011 Compendium of Physical Activities Documenting published MET levels for activities in the 2011 Compendium, based on literature review," vol. 21, no. 2008, pp. 894–899, 2011.
[17] R. Humphrey, "Clinical Applications: The Exercise Caloric Challenge," ACSM s Heal. Fit. J., vol. 10, no. 2, pp. 40–41, 2006.
[18] B. M. Hartmann, C. Schnmidt, and K. Sandfuchs, "Bundeslebensmittelschlüssel (BLS)," 2014.
[19] W. H. Delone and E. R. Mclean, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," J. Manag. Inf. Syst., vol. 19, no. 4, pp. 9–30, 2003.
[20] E. Ammenwerth, S. Woess et al., "Evaluation of an Integrated Telemonitoring Surveillance System in Patients with Coronary Heart Disease," Methods Inf. Med., vol. 54, no. 5, pp. 388–397, 2015.
[21] V. Schusterbauer, "Telehealth-Based Nutrition Management in Support of Diabesity Therapy," Graz, Technical University, 2018.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-322

Is Austria Ready for Telemonitoring? A Readiness Assessment Among Doctors and Patients in the Field of Diabetes

Domenik MUIGG[a,1], Peter KASTNER[b], Robert MODRE-OSPRIAN[b], Daniela HALUZA[c] and Georg DUFTSCHMID[a]
[a] *Center for Medical Statistics, Informatics and Intelligent Systems, Institute of Medical Information Management, Medical University of Vienna, Austria*
[b] *AIT Austrian Institute of Technology GmbH, Graz, Austria*
[c] *Center for Public Health, Institute of Environmental Health, Medical University of Vienna, Vienna, Austria*

Abstract. Background: Telemonitoring offers new opportunities in the treatment of chronically ill patients and could help to improve their quality of life while reducing healthcare costs. Objectives: The willingness to use telemonitoring is examined for both physicians and patients. From the perspective of the most important stakeholders, advantages and disadvantages as well as barriers for telemonitoring are analysed. Methods: A Telehealth Readiness Assessment was carried out with physicians (n=41) and patients (n=47) in a cross-sectional study. A stakeholder survey was conducted by use of interviews (n=28). Results: Average readiness for telemonitoring is 58% for physicians, and 65% for patients. Both are thus in a position where there are several arguments which adversely affect the success of telemonitoring. The most important advantage is the intensified care, while the biggest concerns are data protection as well as the loss of personal communication. The greatest barriers are the lack of funding, the weak clinical and economic evidence and the organisation of the Austrian healthcare system. Conclusion: There are still some barriers to overcome, especially financial, political and organisational.

Keywords. Telemedicine, Managed Care Programs, Austria, Diabetes Mellitus, Surveys and Questionnaires, Attitude of Health Personnel.

1. Introduction

Whether the efforts of research and industry in the field of telemonitoring in recent years have paid off and will now be a success in real life, depends essentially on two questions. Are patients willing to regularly use the sensors and transmit the data? And are doctors ready to evaluate this data and incorporate it into their clinical decisions?

Telemonitoring is the telemedical monitoring of patients who are chronically ill or prematurely released from inpatient treatment in their home environment. In October 2016, a search for "telemonitoring" in the medical literature database PubMed yielded 989 results. Although positive effects of telemonitoring have been demonstrated in several studies, e.g. in [1], there are also those that do not find any significant benefits, such as in [2-5]. However, the American healthcare community seems confident about

[1] Corresponding Author: Domenik Muigg, Center for Medical Statistics, Informatics and Intelligent Systems, Institute of Medical Information Management, Medical University of Vienna, Austria, E-Mail: domenikmuigg@gmail.com.

the merits of telehealth solutions, with one out of two clinics in the US already offering telemonitoring [6].

In principle, the Austrian healthcare environment should also be well prepared for using telemonitoring. Smartphones and Internet access are available for almost everybody [7], the IT skills of the population, including the elderly, are getting better [8]. The number of "digital natives" and "digital immigrants" will naturally increase steadily over the next few years, while the number of "digital outsiders" will continue to decline. In general, the acceptance of modern technologies in the healthcare sector is becoming ever higher, as a survey by Bitkom in Germany [9] makes clear: a total of 6 out of 10 German citizens are open to telemonitoring. Already one third could even imagine using skin-implanted microchips to monitor body function. So why is telemonitoring not yet available in Austria's standard care?

The Austrian healthcare system relies on evidence-based medicine. Several authors identified a need for further studies to determine whether telemonitoring can really improve the quality of life while reducing healthcare costs [3,4,10]. In general, the implementation of telehealth among general practitioners in Europe is not very advanced, with 1.4 out of 4 achievable points, as shown by a survey of the European Commission in 2013 [11].

Indeed, the concept of telemonitoring is not well-known among Austrian patients and physicians, with only 10.5% feeling well or very well informed [12]. However, the participating persons, especially the physicians, are an important part of a successful telemonitoring system [13]. Doctors should therefore be included in the planning of telemedical services [14].

In the present study, we thus aimed at identifying the readiness to use telemonitoring specifically among diabetes patients and physicians treating diabetes patients. We chose diabetes as there is scientific evidence that telemonitoring could be especially useful for this highly prevalent disease [2]. Furthermore, we examined the perceived advantages and disadvantages of telemonitoring as well as possibly existing barriers for the Austrian healthcare system.

2. Methods

Two different methods were chosen for this study. For the question of whether Austrian patients diagnosed with diabetes and the doctors involved in the treatment are ready for telemonitoring, a cross-sectional study by questionnaire was carried out (Readiness Assessment). For the qualitative aspects of the questioning, such as advantages and disadvantages as well as barriers, an expert survey among stakeholders was conducted by use of interviews, as well as various open questions added to the questionnaire of the cross-sectional study.

The survey protocol was approved by the ethical committee of the Medical University of Vienna, Austria, (No. 1197/2017) and conducted following the guidelines of the Declaration of Helsinki.

2.1. Readiness Assessment

2.1.1. Study design

We carried out a cross-sectional study among Austrian diabetes patients and practitioners involved in the treatment. We applied a questionnaire-based scoring system to determine whether the participant was ready for telemonitoring or not [15]. Colloquial the term readiness could be understood as a combination of dissatisfaction with status quo (core readiness), motivation to use telemonitoring to improve the current situation (engagement readiness) and the available resources to accomplish this (structural readiness). The questions were designed in a way that full agreement (4 points) indicated high readiness. The total points scored over all questions indicated whether the participants were in a good position (more than 70% of maximum achievable points), experienced several hindering factors (50% to 70%) or even barriers to use telemonitoring (less than 50%). The assessment tool is designed to be modified in order to meet the country-specific health systems environment. In our study, we slightly adapted the tool. First, we did not offer the option "other" and second, we used a four-point Likert scale, not the original five-point one. The questionnaires were translated from English to German. The German questionnaire is available on request from the authors. We pilot-tested the survey with five laypersons and five researchers experienced in questionnaire design. The data from these tests were not included in the final data analyses. The questionnaire was adapted according to the received feedback.

The online study was open and accessible for practitioners from March 6 to June 2 2017 (89 days) and from April 10 to June 30 2017 (82 days) for patients. We used SoSci Survey (www.soscisurvey.de) as a free, electronic web-based survey tool. The software avoided missing answers, so only complete questionnaires without missing values were available. The survey included a cover letter to inform participants about the scope of the survey and use of the collected data. Since participation was voluntary, consent was implicitly obtained by completing the questionnaire. Once a participant completed the survey, an electronic cookie prevented multiple submissions from the same computer. All responses were anonymous. Data were stored securely and were protected from unauthorized access. We did not offer any incentives for participation.

2.1.2. Study sample and data collection for practitioners

We used a nonrandom purposive sample of German-speaking healthcare experts working in one of the nine Austrian federal states. Private and panel doctors who treat diabetes, mainly general practitioners and internists specializing in endocrinology, were eligible for participation. We identified possible participants through professional networks and associations. Potential participants were recruited based on online email address lists available from national healthcare agencies and the Austrian Diabetes Society which initially invited 863 practitioners. We also contacted representatives of relevant organizations and networks to distribute the survey within their organizations. We sent personalized email invitations containing a link to the questionnaire to these experts as well as reminder emails two and four weeks after the initial contact to prompt further completions. Participants were asked to complete the survey and also forward the survey link to eligible colleagues.

2.1.3. Study sample and data collection for patients

We used a nonrandom purposive sample of German-speaking diabetes patients living in one of the nine Austrian federal states. Potential participants were recruited via various online platforms on which the invitation letter, as well as a link to the questionnaire were included. These platforms were the homepages www.diabetes-austria.com (Diabetes Austria), www.diabetes.or.at (ÖDV), the respective official Facebook pages of both organizations (5072 and 1104 likes), as well as the Facebook Page of "Medtronic Diabetes Austria" (2232 likes). In addition, the invitation was posted to the Facebook groups "Diabetes Type 1 Austria" (1260 members) and "Diabetes Type 2 Austria" (180 members). The Facebook group "Diabetes Kids Austria" (531 members) was not considered, as minors as well as parents of children suffering from diabetes were excluded from study participation.

2.2. Expert Interviews

2.2.1. Study Design

The expert interviews were conducted as direct individual interviews with partially standardized questions (guideline-oriented) and open answers. The interview was recorded, transcribed, anonymised and subsequently evaluated using methods of qualitative content analysis. The participants have been informed immediately prior to the interview about the planned recording, anonymization and retention of the data. Patients gave their consent by accepting the "Participant Information", other stakeholders did so verbally ("Verbal Consent"). The original data of the study was only accessible to the study team.

2.2.2. Study sample and data collection

An initial stakeholder analysis identified potential interview partners. The participating stakeholders were selected randomly or on the basis of personal contacts from the population (network effects). The stakeholders were personally contacted and invited to participate in the interview. In total, 44 individuals or organizations were contacted, whereupon 28 appointments were actually made (participation rate: 64%). Experts were representatives of associations, health politics, existing telemonitoring projects (e.g. “Gesundheitsdialog Diabetes”), research and development, industry and of course users (doctors and patients).

2.3. Statistical analysis

For the questionnaire data, we performed descriptive statistical analyses and present categorical data as absolute frequencies and percentages and continuous data as mean, standard deviation (SD), minimum, maximum and median, where appropriate. We calculated the Spearman correlation coefficient between the metric variable readiness and the ordinal variable attendance. We conducted all statistical analyses using SAS (Copyright © 2017 SAS Institute Inc. Cary, NC, USA).

The interviews were first transcribed and then evaluated by content-structuring qualitative content analysis [16]. MAXQDA11 (VERBI GmbH) was used as QDA software and f4 (dr. dresing & pehl GmbH) as transcription software.

3. Results

3.1. Readiness among practitioners

The first page of the questionnaire was opened 60 times, which corresponds to a response rate of 6.95% regarding the practitioners reached through the available mailing lists. The questionnaire was completed 41 times (response rate with completed questionnaire: 4.75%), which took in average 6:00 minutes (range 2:39 to 10:09 min). The participating practitioners were representative of the Austrian medical profession in terms of age, gender, and federal state. The average age was 51.4 years (md=54, SD 9.4, 30 to 72 years), 56.1% were male.

Participants were predominantly internists (73.2%), while only 12.2% were general practitioners. Above average 51.2% of respondents were private practitioners, 19.5% had a contract with a health insurance institution and 22% worked in a hospital. Further 31.7% of practitioners participated in the nationwide disease management program *"Therapie Aktiv"*. The practitioners think mostly positive about telemonitoring (61%) and were generally open to innovations (65.9% perceived themselves as Innovators or Early Adopters).

The readiness of doctors for telemonitoring was in average 58.2% (95% CI: 53.9 - 62.5). This score thus indicates an average readiness situated in the category between 50% and 70%, according to which several arguments have an unfavorable influence on the use of telemonitoring by practitioners. Moreover 19.5% were in a good position, while 31.7% experienced barriers to use telemonitoring. No significant correlations between readiness for telemonitoring and the demographic characteristics of age, sex or medical specialization were detected.

As shown in table 1 structural readiness is much lower than readiness to participate (engagement readiness) and dissatisfaction with the health system (core readiness). 70.8% of the practitioners are more inclined, 29.3% would definitely offer telemonitoring for their patients (attendance). The Spearman correlation coefficient between readiness and attendance is 0.72.

3.2. Readiness among diabetes patients

The first page of the questionnaire was opened 73 times and the questionnaire was completed 48 times. One data set was excluded from the analysis because the participant did not have diabetes, so that 47 questionnaires were used for the analysis. The full answering took in average 8:33 minutes (range 4:45 to 14:20 min).

The average age was 44.2 years (md=42, SD 15.5, 18 to 77 years), 57.4% of the participants were male. The participating patients were representative of the Austrian population in terms of federal state. The most evident difference of this study population in contrast to all diabetes patients in Austria is the distribution of type 1 diabetics (T1D) and type 2 diabetics (T2D). While in the total population only about 10% are T1D, their portion was 61.7% in our study population. From the 34.1% of T2D, 62.5% were insulin dependent. Some people lived with diabetes for many years (max. 45), but most of them between 0 and 10 years (30%). Diabetes was stable with 59.6%, unstable with 6.4%, and alternating with 31.9%. Self-control was documented electronically by 70.2% of respondents, with only 19.1% still using a paper diary. The patients surveyed think mostly positive about telemonitoring (61.7%) and were generally open to innovations (55.3% perceived themselves as Innovators or Early Adopters).

Table 1. Readiness Assessment among practitioners and patients, showing the absolute achieved points for the three categories. The numbers behind the category names indicate the maximum reachable points in this category (practitioners - patients).

	Practitioners (n=41)					Patients (n=47)				
Category	**Mean**	**SD**	**Median**	**Min**	**Max**	**Mean**	**SD**	**Median**	**Min**	**Max**
Core Readiness (12-20)	8.1	1.8	8	3	12	10.1	3.7	10	5	17
Engagement readiness (28-28)	18.4	4.8	18	8	28	22.9	3.5	24	11	28
Structural readiness (28-20)	13.1	4.4	13	7	24	10.9	2.8	11	6	19
Total (68-68)	39.6	9.1	38	26	60	44	5.1	45	30	53

Patient readiness for telemonitoring averaged 64.6% (95% CI: 62.4 - 66.9). This score thus indicates an average readiness situated in the category between 50% and 70%, according to which several arguments have an unfavorable influence on the use of telemonitoring by patients. Moreover 27.3% were in a good position, while 4.3% experienced barriers to use telemonitoring. No significant correlations between readiness for telemonitoring and the demographic characteristics of age, sex or education were detected.

As shown in table 1, structural readiness and dissatisfaction with the health system (core readiness) are much lower than willingness to participate (engagement readiness). Overall 83% of the patients are more inclined to telemonitoring, 44.7% would definitely attend. The Spearman correlation coefficient between readiness and attendance is 0.43.

3.3. Advantages, disadvantages and barriers

Between February and June 2017, 28 interviews were conducted. The conversations lasted in average 34 minutes (range 12 to 76 min). In total, 116 individuals could be reached in the study.

The most mentioned advantage for telemonitoring was the intensification of care provided by this type of treatment. This was mentioned a total of 40 times, 8 times in expert interviews, 18 times in the questionnaire for patients and 14 times in the questionnaire for doctors (40 | 8/18/14). The second most important advantage was the potential shortening of travel and waiting times (21 | 7/8/6). In the questionnaire for patients we asked how long the consultation with the doctor (on average 18 minutes) and how long the doctor's appointment, including the journey and waiting time, lasted (on average 129 minutes). Through the online processing of a consultation, an average of 111 minutes per appointment could be saved. The third most important advantage, according to respondents, was a better therapy adjustment (21 | 7/7/7).

The three main disadvantages were concerns about data protection (21 | 3/13/5), loss of personal communication and over-emphasis on blood glucose levels (15 | 0/6/9). It is also feared that while telemonitoring will be a huge effort for doctors, the benefits will more likely be on the patient side (8 | 2/0/6).

According to the interviews, the most important barrier in the Austrian healthcare system was the lack of funding (15 times), followed by the lack of high quality long term clinical trials for decision-makers and uncertain economic benefits (11 times). Several stakeholders have therefore emphasized that it is important to establish telemonitoring as part of a structured disease management program, like in HerzMobil Tirol [17]. In seven interviews, the absence of telemonitoring in standard care was attributed to the

organization of the Austrian health system itself. Particularly in the field of diabetes, treatment takes place both in the predominantly federal state financed hospitals and in the health insurance funded outpatient sector. Integrated care therefore requires a combination of inpatient and outpatient areas. Apart from the fact that decision-making in the highly fragmented healthcare system is already difficult, it comes in the case of telemonitoring to incur expenses in the outpatient area (through more intensive care by the resident physician), but to a cost reduction in the inpatient area (by less serious late effects, such as foot amputations). This leads to a shift in funding and political power relations, making decisions even more complicated.

4. Discussion

We used an auto-perception questionnaire that focused on personal experience that could vary depending on the individual. The methodology was slightly adapted, e.g. it was unclear how the question "other" should be formulated. After consultation with the original authors of the study, it was decided to delete the question "Other", as these questions were never answered in the original study. In addition, the five-part Likert scale was replaced by a four-part to avoid neutral answers that are difficult to interpret.

The methodology of the Readiness Assessment for physicians and patients provided quite plausible and new results, although the sample is rather small, especially for the patients and very distorted by the high number of T1D, which was caused by the available channels to reach diabetes patients online. The fact that apparently more T1Ds deal intensively with their illness and also inform themselves and exchange information online is a plausible indication that this group is particularly well suited for telemonitoring (technical affinity and disease awareness). Overall, while the study population is not representative of all Austrian doctors and patients in the field of diabetes, it may be similar to the potential target group for telemonitoring.

The answers to the individual questions of the Readiness Assessment were easy to explain, as well as the scores achieved in the respective categories. When comparing the received readiness level of respondents with the answer to the question whether they would attend telemonitoring, it becomes obvious that, as expected, those with higher level of readiness were more likely to attend. The high correlations between readiness and attendance for both physicians and patients suggest that our methodological approach provided plausible results for telemonitoring readiness among Austrian diabetes patients and the doctors involved in treatment.

One of the strengths of this study is the multi-method approach applied and the in-depth view of the topic, which could be achieved through the many interviews with stakeholders from different fields. Due to the relatively open questions and the focus on the type of stakeholder, it was attempted to obtain a very comprehensive picture of the topic. Thus, for example, if a point polled 40 times, it does not mean that this point is irrelevant to the remaining 76 people. Rather, it means that it was a particularly relevant issue with these 40 people. Several concrete barriers have been identified and there were many interesting requirements, wishes and ideas for a successful implementation. We are currently working on further publications dealing with these aspects in detail, while we focused on the Readiness Assessment in this article.

Although patients have a slightly higher readiness for using telemonitoring than doctors, both are in a position where there are several arguments which adversely affect the success of telemonitoring in Austria. This is inline with the results from expert

interviews, finding that there are still some barriers to overcome, especially financial, political, organisational and technical.

The authors are grateful to the participants of the pretest for their helpful and valuable detailed comments and suggestions. The authors also express their sincere appreciation to all survey participants.

References

[1] Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008; 14(10):1118–26.

[2] Wootton R. Twenty years of telemedicine in chronic disease management--an evidence synthesis. J Telemed Telecare 2012; 18(4):211–20.

[3] Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K et al. Personalized Telehealth in the Future: A Global Research Agenda. J Med Internet Res 2016; 18(3):e53.

[4] Ekeland AG, Grottland A. Assessment of MAST in European patient-centered telemedicine pilots. Int J Technol Assess Health Care 2015; 31(5):304–11.

[5] Sorknaes AD, Bech M, Madsen H, Titlestad IL, Hounsgaard L, Hansen-Nord M et al. The effect of real-time teleconsultations between hospital-based nurses and patients with severe COPD discharged after an exacerbation. J Telemed Telecare 2013; 19(8):466–74.

[6] Marcoux RM, Vogenberg FR. Telehealth: Applications From a Legal and Regulatory Perspective. P&T 2016; 41(9):567–70.

[7] Statistik Austria. IKT-Einsatz in Haushalten 2015; 2016 Aug 8 [cited 2016 Oct 13]. Available from: URL: https://www.statistik.at/web_de/statistiken/energie_umwelt_innovation_mobilitaet/informationsgesellschaft/ikt-einsatz_in_haushalten/index.html.

[8] MindTake Research GmbH. Mobile Communications Report – MMA 2016. Wien; 2016 Aug 18 [cited 2016 Oct 13]. Available from: URL: http://www.mmaaustria.at/.

[9] Rohleder B. Digital Health. Berlin: Bitkom; 2016 Sep 15 [cited 2016 Oct 13]. Available from: URL: https://www.bitkom.org/Presse/Anhaenge-an-PIs/2016/Bitkom-Pressekonferenz-Digital-Health-15-09-2016-Praesentation-final.pdf.

[10] Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012; 81(1):1–11.

[11] Codagnone C, Lupiañez-Villanueva F. Benchmarking deployment of eHealth among general practitioners (2013): Final report. Luxembourg: Publications Office; 2013.

[12] Haluza D, Naszay M, Stockinger A, Jungwirth D. Prevailing Opinions on Connected Health in Austria: Results from an Online Survey. Int J Environ Res Public Health 2016; 13(8).

[13] Pecina JL, Vickers KS, Finnie DM, Hathaway JC, Takahashi PY, Hanson GJ. Health Care Providers Style May Impact Acceptance of Telemonitoring. Home Health Care Management & Practice 2012; 24(6):276–82.

[14] Hannan TJ, Celia C. Are doctors the structural weakness in the e-health building? Intern Med J 2013; 43(10):1155–64.

[15] Legare E, Vincent C, Lehoux P, Anderson D, Kairy D, Gagnon M-P et al. Developing and validating the French-Canadian version of the practitioner and organizational telehealth readiness assessment tools. J Telemed Telecare 2010; 16(3):140–6.

[16] Kuckartz U. Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. Weinheim: Beltz-Juventa; 2012.

[17] Von der Heidt A, Ammenwerth E, Bauer K, Fetz B, Fluckinger T, Gassner A et al. HerzMobil Tirol network: Rationale for and design of a collaborative heart failure disease management program in Austria. Wien Klin Wochenschr 2014; 126(21-22):734–41.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-330

Challenges of a HL7 CDA Guideline for Telehealth Based DMP Systems

Nikola TANJGA[a,b,1], René BARANYI[b], Thomas GRECHENIG[b], Christian GOSSY[a], Stefan WELTE[a], Peter KASTNER[a] and Robert MODRE-OSPRIAN[a]
[a]*AIT Austrian Institute of Technology, Vienna, Austria*
[b]*Research Group for Industrial Software (INSO), TU Wien (TUW), Vienna, Austria*

Abstract. Background: Disease management programs (DMP) are a modern way of treating health conditions and are becoming a part of standard care. One telehealth DMP service has been in regular operation since 2017, named "HerzMobil Tirol". Objectives: This paper investigates, if the electronic health record standard HL7 CDA, which is widely accepted in the health care industry, could be used for telehealth DMP services as well. It is already in use in a legally required integrated element of healthcare in Austria called ELGA. An official guideline from the Austrian Ministry of Health sets the standard for telemonitoring with data logging. Methods: After the background knowledge was built up, requirements have been gathered through existing official guidelines and interviews and existing documentation by "HerzMobil Tirol". Results: Twenty-five requirements were collected, categorized and analyzed to determine if the existing CDA guidelines are suitable or a new standard must be designed. Conclusion: Based on the requirements, it was established that seven specific sections and two different CDA documents are needed.

Keywords. Disease Management, Health Level Seven, Electronic Health Records, Computerized Medical Record System, Telemedicine, Patient Generated Health Data

1. Introduction

Within disease management programs (DMP) healthcare professionals (HP) interact and communicate with patients in a coordinated way, to treat a condition over a defined period of time where patient self-care is also highly essential for the therapy [1]. Such DMPs can also be supported by electronic medical devices and webservices, to make a remote medical treatment possible [2].

One of such telehealth services has been in regular operation since 2017 in the Austrian state of Tyrol and is named "HerzMobil Tirol". HerzMobil Tirol is a multidimensional post-discharge disease management program for Heart Failure (HF) patients employing a telemedical monitoring system incorporated in a comprehensive network. This network includes specialized HF nurses, private practice physicians and primary and tertiary referral centers [3]. The HerzMobil Tirol telehealth services collect vital signs on a regular basis like weight, blood-pressure, heart frequency, blood sugar levels, steps and wellbeing through questionnaires, to name a few. They also collect medication intake data, used medical devices and written communication [4]. There is

[1] Corresponding Author: Nikola TANJGA, Research Group for Industrial Software (INSO), TU Wien (TUW), Vienna, Austria, AIT Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria, E-Mail: nikola.tanjga@live.at

no way to get access to the data outside of the HerzMobil Tirol network. Due to the fact of the scientific context it grew out of, and also the possibility of accessing intern information, the HerzMobil system therefore is taken as the reference system for Austrian telehealth DMPs.

The Ministry of Health Affairs in Austria has published a general IT-infrastructure guideline for telemonitoring with data logging [5]. It is describing, how a standardized base for the interfaces between the components could connect different DMP modules and make use of the synergy. For example, patients with more than one chronic disease. It is also mentioning the HL7 CDA Personal Healthcare Monitoring Report (PHMR) [6], that is created with the goal to contain monitoring information on personal healthcare, specifically patient measurement data taken by consumer medical devices. This PHMR is in use in an ongoing telemedicine project in Denmark [7].

ELGA ("ELektronische GesundheitsAkte") in Austria is on its way, to create a connection between patient record systems, that are mandatory for every health care provider. Specific documents from different healthcare professionals currently being registered on a centralized server and the actual data is hosted by the health care provider in their own ELGA domain. The patient can access all uploaded documents from a web based application. Other HPs can get access to the data of all uploaded documents concerning the patient by inserting the patients social security card. All documents in the ELGA must use the standardized document profile HL7 Clinical Document Architecture (CDA) [8].

In this paper the authors address the following questions: Is it possible to make the telehealth DMP data accessible for HP outside the DMP via standardized documents? Will the requirements of the main stakeholders from a reference telehealth DMP fit the given CDA Standard? Is it also possible to make this DMP document comply with the requirements of the ELGA and the Ministry of Health Affairs?

2. Methods

First the general background knowledge was built up which was summarized in the introduction. From this knowledge it was clear which different stakeholders had to be interviewed for further analysis. Before the needed interviews with telehealth service experts, individual recaps and interview guides, that match their profession, were created. The recaps were given a brief overview over the situation and were sent before the interview to the interviewee. The interview started with a short presentation of the recap followed by questions from the interview guide being asked. During questioning answers were written down and repeated to the interviewee, so that a correction from the interviewee was possible.

Some requirements were collected through analysis of the interviews with the stakeholders from the reference telehealth DMP service. The three main stakeholders regarding a telehealth DMP service, as seen on figure 1 in the orange box, include: the experts who are also the system designers in the collaborative heart-failure network, the healthcare professionals that are using the DMP inside the network and the healthcare professionals outside the DMP network, that will read the document. Each group offered one stakeholder representative to be interviewed. The requirements were extracted from the given answers. Additionally, the existing internal documentation of the design team was analyzed and the requirements relevant to the DMP document were collected as well.

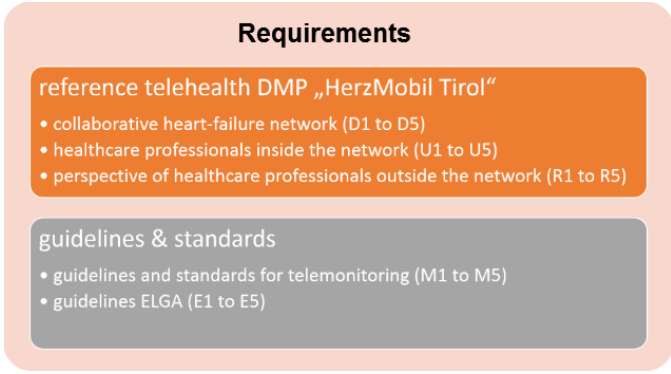


Figure 1. The stakeholder requirements are in two main groups, as shown. The identifiers D1 to D5, U1 to U5 and so on are analogues to the later defined requirements and their IDs.

Beside the DMP stakeholders, there are also two organizations with official guidelines that must be considered, as seen in the grey box on figure 1. From the Ministry of Health Affairs there was published a general IT-infrastructure guideline for telemonitoring with data logging [5]. Inside of this a guideline of HL7 International was referenced [6], that is affecting this work. The ELGA has CDA guidelines already in use for Austrian purposes that need to be considered also [8]. The official guidelines from these two organizations were analyzed and requirements were extracted. The purpose of the found guidelines for standardized records has been compared with the targets of this work.

After the requirements were collected, a categorization was made regarding the depth of the impact of the requirement on the guideline for the DMP CDA record. Because some requirements were targeting the same topic, a grouping inside the categories was also made. By the end of the results every requirement group was being considered and finally important aspects were suggested for a DMP document.

3. Results

In 3.1 the requirements are being listed. In 3.2 the requirements are being divided into categories. In 3.3 the possible solutions are being described. In 3.4 the potential guideline for telehealth DMPs is being summarized.

3.1. Requirements

In the following table, the requirements have been listed and given an ID. The letter in the ID stands for a stakeholder, see figure 1.

3.2. Categorization

In the following section, the requirements are categorized and grouped as seen in figure 2. Three categories were established. “General” requirements are requirements that are affecting the whole document. The requirements in the category “sections” are influencing the individual sections that will include the information in the document. The

Table 1. The individual requirements with their ID

ID	Requirement
D1	For all used data exports, imports and interfaces, only known standards out of the common practice should be used.
D2	There should be a summary of the DMP after it is finished, with all the relevant data for the treatment validated and signed by the supervising physician.
D3	The new document standard should be used as a more detailed and standardized backup file for the HP, to still have access to the data of their patient after the telehealth DMP service has been shut down or if the HP is not active in the DMP network.
D4	The document should be an easy way to pass on the data when the patient or another entitled person like a lawyer of the patient is demanding it.
D5	There should be a way to add diagrams in a widely-used picture-format to the document. Normally one diagram per section should be enough, but the chance of having more than one should be considered.
U1	There are types of comments and notes that are only meant for the HP themselves or among each other, these should not be included in that form in any official document.
U2	Attention should be paid, that for every chronic condition, the common practice is different and changes over time. This is primarily affecting the needed monitoring data.
U3	A fundamental question at the beginning of the document should show a brief overview, what condition caused to start this treatment. This is essential when more than one condition is known and this must be kept in mind while treating the patient.
U4	When written text is required, there should be an option to write various text elements, which should be reusable and combinable.
U5	Laboratory results and other documents should be able to be added to a DMP document.
M1	In the long-term the telemonitoring with data logging guideline will create a homogenous, communicating IT system landscape, that supports telemonitoring for health applications. It will also create a base for the international accordance of these IT standards.
M2	Specifying and further developing of standards for the IT architecture for telemonitoring and making the availability of compliant devices on the market better is also defined as a target in their guideline.
M3	Another goal is the creation of a technically binding base for patient care, especially for patients with multimorbidity, to add medical devices independently from the manufacturer for any requirements and course of diseases. This will be specified in the spirit of a rough concept, that will serve as a base for designing a detailed concept in ongoing planning phases.
M4	They are also specifically defining non-objectives. One worth mentioning for this work is, not defining quality standards or recommendations for the medical aspects of disease management programs, knowing that the existence of a DMP is essential for telemonitoring.
M5	For the data transfer between various interfaces, the guideline is specifically mentioning that the usage of HL7 CDA analogous to the HL7 CDA PHMR is binding.
E1	In the official law in the “ELGA-Verordnung 2015” it is defined that starting from 1.1.2018 the main documents can only be uploaded in “EIS Full Support”, that is analogues to HL7 CDA Level 3, which provides full machine readability.
E2	The data in the CDA header must be consistent to all specifications for headers of the general ELGA CDA guideline “Allgemeiner Implementierungsleitfaden” in its version 2.06.2 [8]. This also includes the correct referencing of the patient and the defaults for automatically generated documents.
E3	If the vital sign components (section, group and entry) are used, it should be considered that the ELGA is already specifying these vital sign components in the ELGA discharge summary [8] for the measurements recorded during an inpatient treatment.
E4	There are already defined code-systems and value sets in other ELGA document standards. These should be reused. If an expansion is needed, the request process has to be investigated.
E5	If medication components (sections, groups or entries) are needed, these should be reused from other ELGA guidelines like from the “eMedikation” [8], a standard for electronic medication CDA documents.
R1	Values outside of a range, that was set by the network physician, can be much more interesting than the standard values. These should be somehow summarized or highlighted.
R2	All the data in the document should be tagged with a timestamp. This is especially important in a DMP document because a treatment in HerzMobil Tirol normally last for 3 months or more.
R3	The additional information of the used measurement device is useful but should not be mandatory.
R4	The chronological order should be ascending, meaning that the oldest value is also the first one on the list.
R5	The details of a contact person should be included in every document, who can answer questions regarding the treatment of the specific patient.

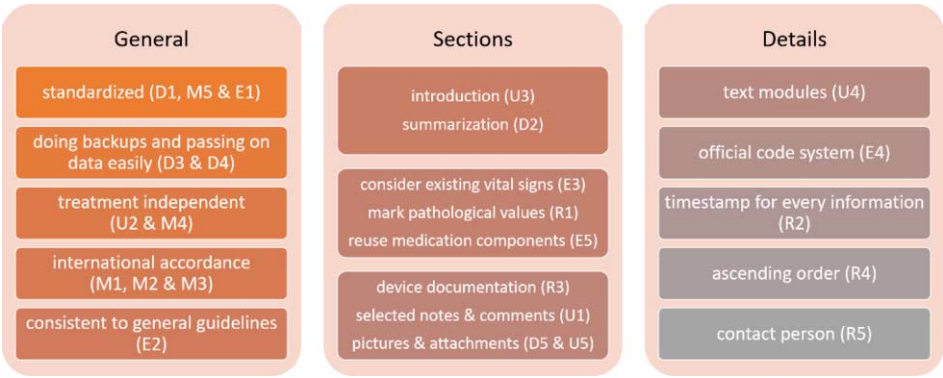


Figure 2. The requirements categorized and grouped by topics.

requirements in the category "details" will only change specifics in the record. Requirements with equal topics were discussed together under one title, followed by the requirement IDs in brackets. Requirements were grouped when the topics were similar but not the same, as seen in the first block in "sections" in figure 2.

3.3. Consideration of every requirement

The data standards for D1 are already limited by E1 and M5. This work has to use the widely-known document standard CDA from HL7 CDA Level 3, meaning that information is fully machine readable. M5 is even restricting it to a document that is analogous to the existing HL7 CDA PHMR standard. If designing a new standard is really needed it is beneficial to compare it to a similar existing standard.

D3 and D4 are both accomplished, because the HL7 CDA Standard is based on single XML document for every treatment. It can even be displayed in a simple browser, when an official XSL stylesheet has been added. Patients have the possibility to get the documents that were created out of the DMP directly in their ELGA portal.

U2 is stating, that different diseases are demanding different treatments and that treatments are evolving over time. The official document from the Ministry of Health Affairs is stating with M4, that they don't want to make a standard for the DMP treatment itself, so that a doctor has to follow a specific standard operation procedure when a disease is diagnosed. This is also not the objective for this work.

M1 is satisfied, providing that the HL7 CDA standard is an international accepted standard, used in many variations over the world. The more specific HL7 CDA PHMR standard is designed for medical devices generating a HL7 CDA document and sending it directly to the patient record folder. This fact is supporting M2 and M3, so that the devices could speak a common document export language and be used in different DMPs.

E2 demands that the document is consistent to the general ELGA CDA guideline. The interesting aspect is, that the standard should be like the HL7 CDA PHMR, which is automatically generated without any user interaction. Such a document is not having a human validation, that is required in most of the ELGA documents. Also required is the versioning of the ELGA document. It could add values to an existing document with every measurement, creating a new version number and extending the treatment-timespan of the document.

Regarding the sections U3 clarifies that a brief introduction into the DMP document is needed. Normally, a title can be insightful and clearly state the targeted disease. However, the possible condition with more than one chronic disease, which are influencing each other, demands to be stated first in documents like these. It would be too much information to be stated in a simple title. D2 is asking for a summarization of the treatment. This could be written continuously while treating or once at the end. A document could be used that is similar to the ELGA discharge summary "Ärztlicher Entlassungsbrief", that is meant as a discharge summary after an inpatient treatment. It could be a signed document and would include the data of the DMP.

E3 requires that the vital signs section in the PHMR is adapted to the vital signs section in the general ELGA guideline, that is demanding that every parameter in the vital signs is specified with their official name, unit and code in the value set "ELGA_Vitalparameter". The PHMR is also defining, that the vital signs are specific parameters, that are been standardized in health care. Non standardized parameters are being collected in the results section. R1 is demanding that parameters that are out of a specific range, get highlighted or summarized somehow. The HL7 CDA standard itself is limiting the formatting possibilities. The E2 with its general ELGA CDA guideline is adding more possibilities for formatting with extended style-codes, that a ELGA CDA stylesheet can realize. Aside from headings and table formatting, they also added xELGA_blue for highlighting text sections and xELGA_red for highlighting a whole row of pathological data of laboratory measurements. Aside of formatting, the table can be made collapsible with the given stylesheet, summarizing only the pathological or out-of-range data at first. With a click the table can expand and show all the data. E5 is similar to the usage of medication components that the ELGA discharge summary "Ärztlicher Entlassungsbrief" is using. The one big difference for the DMP document standard is, that it also needs to record the intake of the medication. This should be a machine-readable entry itself. Because such a medication intake entry is not existing in the ELGA guidelines at the time of this work, one has to be created, similar to the existing ELGA medication components.

R3 is weakening the mandatory medical equipment section in the PHMR from M5, because its stating that the added medical equipment information is not of high value. The design-purpose of a PHMR is that a device itself would create such a document. Because of this, the medical equipment section in the PHMR should be filled with the information of the device itself. The purpose of writing a document out of a DMP system whereby one or more devices are delivering vital parameters is making this section optional, to not overcomplicate the implementation of such a document. U1 could be simply solved through selecting the notes and comments at the exact moment a document is generated. Also, the option of categorization of comments and notes beforehand would be a practical solution. That would support the automatic generation of the document without any user interaction. Just the categories of notes and comments that can be uploaded needs to be set beforehand. D5 is requiring an option to add pictures to a section. The general ELGA CDA guideline is offering "observationMedia"-entries. These are embedding one or more attachments in a CDA as a base64 coded block of text to a section directly. Defined in the "ELGA_Medientyp" valueset there are PDF, MPEG, XML and the appropriate image types GIF, JPEG and PNG. Within this requirement, U5 also got answered, which demands the option to add whole documents as XML or PDF to a CDA.

The detail requirement U4 is not directly influencing the document standard. The text modules could offer one big text block or single paragraphs.

Table 2. recommended CDA sections for a DMP record

name	explanation
fundamental question	As a mandatory section the fundamental question could briefly show details about the reasons for starting the DMP treatment with adjustments on the patient.
summary	The summary outlines the whole DMP treatment. Every time a significant observation or change in treatment has been made, a few sentences should be added.
vital signs	The vital signs section is the observation section for clinical defined parameters. Like in the HL7 CDA PHMR, the DMP document should require both observation sections or at least one of these two to be valid.
results	The results section is the other observation section for non-clinical defined parameters. The medication intake can be documented in the result section as well. For this a CDA entry must be specified, so that that data is standardized for machine readability.
medical devices	The medical devices, unlike in the PHMR, should be optional, but when used it should support full machine readability.
feedback	The feedback contains all the notes and comments that were selected to share with HP outside of the DMP network.
attachment	In the attachment, whole documents like laboratory results (e.g. as PDF) can be appended.

E4 is saying that codeable parameters should be machine readable through a code system. There is the possibility that even in the good cultivated database there may not be a parameter that could be measured by a medical device. For those rare cases a publicly available value set from the HP or medical device manufacturer themselves could be created or updated with these missing values.

R2 is a point that could be specified as a rule in the specific standard guideline, that every data, even simple text passage, should have a timestamp or at least the date.

R4, with its ascending chronological order, should be standardized in the guideline. It would ensure that every document has been written and can be read the same way.

R5 is even matching to the CDA HL7 standard, that gives every document a contact person, who can be asked regarding questions to the document. This should be decided for every DMP separately, but at least one contact person should be given.

3.4. Suggested guideline for telehealth DMPs

From the above collected and discussed requirements, a guideline for a DMP record can be proposed. A standardized document for DMPs could have, besides fulfilling the header requirements of the overall ELGA CDA guideline, seven different CDA sections in their body, presented in table 2.

There needs to be distinction between the two documents. One is being created while treatment is ongoing. This could be automatically generated and therefore not be specifically signed. The second document is summarizing the whole treatment at the end of a DMP, signed by the supervising physician. Two different CDA subclass standards would be useful. The document that is used during treatment would contain the above specified sections. This could be created by the telehealth DMP system automatically on a daily basis. It could be called DMP report or short DMPR. The discharging document could be a subclass of the official ELGA discharge summary. In addition of being compliant to the ELGA discharge summary, it could contain the last DMP report itself as an attachment or include all the sections of the last report solely. This document type could be called DMP discharge summary, short DMPDS.

4. Discussion

The requirement analysis is showing, that the HL7 CDA is the right match for a base standard for such a DMP document standard. The HL7 CDA PHMR could be used as a template to create an ELGA guideline for the DMPR that is compliant to the general CDA guideline of the ELGA. Besides this, a DMPDS must be specified and that could be an adjusted version for the DMP of the actual ELGA discharge summary. It should be investigated if a third DMP document type is needed, however it is not specified or asked in the collected requirements. It would summarize a DMP treatment and be signed by a physician while the treatment is ongoing. This could be called a DMP situation summary or a DMP situation report.

In the future, the requirement for writing selectable text modules could be the reason for bringing up a code system for whole text blocks. Therefore, the meaning of the text block could be coded. For further discussion on the requirement to add diagrams to the document, SVG should be considered. This is due to the fact that most of today's browsers are compatible with such vector graphics. They would show its advantages in small storage spacing and enlarging without getting grainy, especially for diagrams. Additionally, ongoing work from other countries, like the discussion in Denmark about the addition of methodCodes for every measurement [7], should be considered in the ultimate implementation guideline for DMP documents.
It also should be noted, that this requirement analysis, made with the tools described, could be incomplete. All of the found requirements and their analysis stated in this work should be an appropriate start for designing the two needed CDA documents that keep record on a DMP, especially on a telemonitoring supported DMP.

Acknowledgements

The authors thank Momen Radi for sharing his wisdom with us.

References

[1] Moertl, Deddo, et al. Disease management programs in chronic heart failure. Wiener klinische Wochenschrift 129.23-24 2017. pp. 869-878.
[2] Demiris, George, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. Journal of the American Medical Informatics Association 15.1, 2008. pp. 8-13.
[3] Pölzl, Gerhard, et al. HerzMobil Tirol: a collaborative heart failure disease management program in Austria. Spitzenforschung Herz-Kreislauf-Medizin 2016, 2016. pp. 94 - 95.
[4] Von der Heidt, Andreas, et al. HerzMobil Tirol network: rationale for and design of a collaborative heart failure disease management program in Austria, in: Wiener klinische Wochenschrift 126.21-22, 2014. pp. 734-741.
[5] Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz, Rahmenrichtlinie IT-Infrastruktur-Telemonitoring Messdatenerfassung, https://www.bmgf.gv.at/home/Rahmenrichtlinie_IT-Infrastruktur-Telemonitoring_Messdatenerfassung, last access: 21.1.2018
[6] Health Level Seven International, Hl7 Standards Product Brief - HL7 CDA® R2 Implementation Guide: Personal Healthcare Monitoring Report, Release 1 , http://www.hl7.org/implement/standards/product_brief.cfm?product_id=33, last access: 1.2.2018
[7] Christensen, H. B., Localizing the HL7 Personal Health Monitoring Record for Danish Telemedicine, in: MIE, The Swedish Exhibition & Congress Centre, Gothenburg, 2014. pp. 251-255.
[8] ELGA GmbH, ELGA: Startseite, https://www.elga.gv.at/index.html, last access: 21.1.2018

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

doi:10.3233/978-1-61499-858-7-338

Achieving Interoperability Between Arden-Syntax-Based Clinical Decision Support and openEHR-Based Data Systems

Martin SPINETH[a,b], Andrea RAPPELSBERGER[c] and Klaus-Peter ADLASSNIG[b,c 1]
[a]*University of Applied Sciences Technikum Wien, Vienna, Austria*
[b]*Medexter Healthcare GmbH, Vienna, Austria*
[c]*Section for Artificial Intelligence and Decision Support, Medical University of Vienna, Austria*

Abstract. Background: Arden Syntax is an international standard for medical knowledge representation and processing. The openEHR Foundation publishes specifications for the creation of electronic health records based on interoperable data models. Objectives: To perform a feasibility study showing how Arden Syntax medical logic modules (MLMs) can access openEHR data. Methods: Medexter's ArdenSuite was applied as an implementation of an Arden-Syntax-based clinical decision support framework, and Marand's EhrScape as an implementation of an openEHR system, for the purpose of data exchange. To assess their interoperability, we developed a use case in which ArdenSuite was connected to EhrScape; the purpose was to determine whether a patient suffers from orthostatic hypotension based on data supplied by EhrScape and decision support provided by Arden Syntax MLMs. Results: An archetype query language request was sent from an MLM to EhrScape, and the results were sent back. Conclusion: This preliminary study clearly shows that the ArdenSuite's MLMs can communicate with openEHR-based data sources.

Keywords. Health Level Seven; Electronic Health Records; Decision Support Systems, Clinical; Health Information Interoperability; Hypotension, Orthostatic.

1. Introduction

1.1. Objectives

Since a connection between the Arden Syntax and openEHR-based systems has not been reported yet in the published literature, the aim of this work was to investigate their combination and report on the experiences.

ArdenSuite [1] is a commercial solution that allows writing, compiling, and uploading Arden Syntax medical logic modules (MLMs) for clinical decision support (CDS) as well as MLM execution and integration with medical host systems and data sources. EhrScape [2] is a cloud-based openEHR platform that was used as a representative of an openEHR-based database system.

To test interoperability we established a small Arden Syntax knowledge base consisting of two MLMs. The knowledge base is used to determine whether a patient has

[1] Corresponding Author: Univ.-Prof. DI Dr. Klaus-Peter Adlassnig, Medexter Healthcare GmbH, Borschkegasse 7/5, 1090 Vienna, Austria , E-Mail: kpa@medexter.com.

orthostatic hypotension, based on blood pressure measurements. The necessary data were taken from EhrScape.

1.2. Arden Syntax

Arden Syntax [3] is a Health Level Seven (HL7) International [4] standard for the representation and processing of medical knowledge. This standard allows knowledge to be represented in a computer-executable format and can therefore be used for the development of CDS systems [5]. Currently, HL7 International's Arden Syntax Workgroup [6] fosters the development of the Arden Syntax standard.

So-called MLMs constitute the basic knowledge elements of Arden Syntax. Each MLM contains sufficient logic to support at least one medical decision. An MLM can receive data either from input arguments, other MLMs, or via so-called read or curly brace statements from external data sources, such as an electronic health record (EHR). Figure 1 shows a code snippet of an MLM's knowledge section, and Figure 2 the corresponding medical explanation.

One advantage of Arden Syntax is that it resembles natural language (see example in Figure 1). This makes MLMs easy to write and read, even for persons with less extensive programming skills [5]. In addition, using Arden Syntax makes it easier for hospitals to share CDS knowledge packages, because they are independent of the programming language and other technical specifications used in the respective hospital's information system. Furthermore, backwards compatibility was preserved over the Arden Syntax versions [7]. As a result, knowledge packages written in an older Arden Syntax version can be used in systems with the latest Arden Syntax version, which is version 2.10 [3].

```
knowledge:
    type: data_driven;;
    data:
        LET getBloodPressure BE MLM 'getBpEhrScape' FROM Institution "Medexter Healthcare GmbH";
        include getBloodPressure;
        ;;
    priority: ;;
    evoke: ;;
    logic:
        LET bloodPressure BE call getBloodPressure;

        IF TIME OF bloodPressure[2] IS WITHIN 180 SECONDS FOLLOWING TIME OF bloodPressure[1] THEN
            LET firstBp BE bloodPressure[1];
            LET secondBp BE bloodPressure[2];
        ELSEIF TIME OF bloodPressure[1] IS WITHIN 180 SECONDS FOLLOWING TIME OF bloodPressure[2] THEN
            LET firstBp BE bloodPressure[2];
            LET secondBp BE bloodPressure[1];
        ELSE
            LET notification BE LOCALIZED 'Error';
            conclude true;
        ENDIF;

        IF (firstBp.systolic - secondBp.systolic) IS GREATER THAN OR EQUAL 20 OR
           (firstBp.diastolic - secondBp.diastolic) IS GREATER THAN OR EQUAL 10 THEN
            LET notification BE LOCALIZED 'OrthostaticHypotension';
            conclude true;
        ENDIF;
        ;;
    action:
        return notification;
        ;;
    urgency: ;;
```

Figure 1: Example snippet of the knowledge section of an MLM for orthostatic hypotension notification

Orthostatic Hypotension Notification

ALERT
if
Time of 2nd blood pressure measurement is within 3 min after the time of the 1st measurement
and
Systolic value of 2nd blood pressure measurement is at least 20 mmHg lower than systolic value of the 1st
and/or
Diastolic value of 2nd blood pressure measurement is at least 10 mmHg lower than diastolic value of the 1st

Figure 2: Medical explanation of the orthostatic hypotension notification MLM (see Figure 1)

1.3. OpenEHR

OpenEHR [8] is an open standard specification with the goal of turning physical health data into an interoperable digital form. To achieve this goal, the openEHR Foundation publishes specifications for the development of EHRs [9]. The openEHR approach works with multi-level, single-source modeling. 'Multi-level' means that domain experts develop models for the semantics of clinical information systems (archetypes) in their own layer, separated from the technical definitions [10]. 'Single source' means that archetypes and templates are developed independent of specific document or messaging standards. As a result, specific models (such as *microbiology results*) only need to be modeled once in order to generate reports, documents, user interface forms, representational state transfer (REST) application programming interfaces (API) specifications, or other message formats [10].

The archetype query language (AQL) [11] was specifically designed by openEHR for searching and retrieving data from archetype-based EHRs. Figure 3 shows an example of an AQL request inside a curly brace expression. Its advantage over other query languages such as SQL is that it is independent of a specific data model implementation. This is possible because the queries are expressed at the archetype (semantic) level and not at the data instance level. The minimum requirement for AQL to work is that the data is marked with the appropriate archetype codes and terminology codes. AQL queries can be sent to the openEHR server via different types of interfaces, including REST interfaces generated with openEHR templates. The openEHR specification requires that services support at least XML or JSON for data representation [12]. (However, this openEHR specification is still under development per 30 January 2018.)

```
19    data:
20        measurements := READ
21        { openEHR:query/?aql=
22            select
23               bp/data[at0001|history|]/events[at0006|any event|]/time as time,
24               bp/data[at0001|history|]/events[at0006|any event|]/data[at0003]/
25                   items[at0004|Systolic|]/value/magnitude as systolic,
26               bp/data[at0001|history|]/events[at0006|any event|]/data[at0003]/
27                   items[at0005|Diastolic|]/value/magnitude as diastolic
28            from EHR e
29            contains OBSERVATION bp[openEHR-EHR-OBSERVATION.
30                   blood_pressure.v1|Blood_Pressure|]
31            where e/ehr_id/value='52a6f911-9bf2-4de0-9513-64e5a6798c5a'
32            order by time desc
33            limit 2
34        };
```

Figure 3: AQL query within a curly brace expression of an Arden Syntax MLM that is used to request two blood pressure measurements from a specific patient in EhrScape

1.4. Orthostatic hypotension

Orthostatic hypotension [13] is a cardiovascular disorder whose prevalence in the population increases with growing age. It is diagnosed when the blood pressure of a patient drops by at least 20 mmHg systolic or 10 mmHg diastolic within 30 to 180 seconds after standing up. Symptoms include nausea, fatigue, light-headedness, dizziness, "coat-hanger" pain, visual blurring, and syncope [13]. Since the majority of patients are asymptomatic or have just a few nonspecific symptoms, a large number of cases remain undetected.

2. Methods

2.1. ArdenSuite

ArdenSuite [1] is a CDS technology platform based on the Arden Syntax standard and was developed by Medexter Healthcare [14]. The ArdenSuite integrated development and test environment (IDE) allows users to write and compile Arden Syntax MLMs. ArdenSuite supports all ArdenSyntax versions. Subsequently, MLMs can be tested within the integrated test environment. The ArdenSuite server enables access to client applications. The compiled MLMs can be uploaded to and managed by the ArdenSuite server. Via a REST or SOAP interface, external applications can call the deployed MLMs and retrieve the results. Using the ArdenSuite connectors and extensions, the ArdenSuite server can connect to different data sources such as "normal" databases (e.g., SQL) or a fast healthcare interoperability resources (FHIR) server. Additionally, there is an Activiti extension that connects ArdenSuite to the Activiti Business Process Model and Notation (BPMN) platform [15].

In this project, the ArdenSuite IDE was used for the development of MLMs to test the connection of Arden Syntax to openEHR. Besides, the Arden Syntax server in combination with the FHIR Connector were used to establish a connection with an openEHR system and execute the MLMs.

2.2. EhrScape

Marand d.o.o. company [16] developed the *Think! EHR Platform*, which is an openEHR-based EHR system. EhrScape offers an open API/open data version of this platform, which is accessible to developers in the cloud. It includes documented REST APIs [17].

Users have to register on the webpage in order to access the platform. The EhrExplorer [18] can be accessed by entering the respective account details. This is a user interface for administrating the EHR components. Managing archetypes, templates, and queries as well as executing queries are some of the functions of the EhrExplorer. The login data are also needed for the authentication of REST requests from outside the EhrExplorer.

After registration for an EhrScape access, the provided test data were searched for suitable datasets for the application of our use case (Table 1). The EhrExplorer was used to formulate a query to search the test data for patients that include the blood pressure archetype and provide a sufficient number of measurements. Furthermore, the connection from an external system and the reception of data were tested by using the freeware Postman [19] to send REST requests to EhrScape.

Table 1: Use case description for orthostatic hypotension

Title	Orthostatic Hypotension
Actors	• Physician • OpenEHR-based EHR • ArdenSuite server
Prerequisites	• An openEHR-based system that can store blood pressure measurements • The ArdenSuite server with the appropriate connectors or extensions, configured to connect to the EHR • An MLM that contains the logic for determining whether a patient suffers from orthostatic hypotension, based on blood pressure measurements
Steps	**1.** With the intention of testing the patient for orthostatic hypotension, the physician determines multiple blood pressure values while performing the bedside active standing test or the head-up tilting test [20]. **2.** The blood pressure values are entered either manually or automatically into the patient's EHR. **3.** Activated by a data-based event or a direct call by the physician, the MLM "Orthostatic Hypotension" is called from the ArdenSuite server. **4.** After successful execution of the MLM, the ArdenSuite server returns the computed conclusion.

2.3. Use case description

We developed the use case to assess the interoperability between ArdenSuite and EhrScape by retrieving and processing openEHR data (see Table 1). We created a possible workflow for orthostatic hypotension by using Arden Syntax MLMs that retrieve data from an openEHR database. Successful execution of steps 3 and 4 of this use case would indicate that the Arden Suite can connect to openEHR systems and process their data.

In step 3 of the use case, after the MLM is called the ArdenSuite server should retrieve the blood pressure measurement values from the openEHR server. After receiving these values, the MLM is executed and a conclusion is drawn according to the logic in the MLM. Step 4 serves as confirmation that the data were received and processed, and the results returned.

3. Results

After analyzing EhrScape's API explorer and testing the connections with Postman, the "/query" REST interface was chosen. This interface accepts an AQL expression as a parameter that is executed; the result is returned in the JSON or XML format. We selected the XML representation of data.

A review of the provided test data in the EhrExplorer led to the conclusion that there is enough data for testing, but additional test datasets need to be injected into the database. A patient with 27 blood pressure measurements was found. Two MLMs were written for testing the interoperability of the ArdenSuite with EhrScape and realizing the use case. The first MLM sends the data request to the EhrScape server via a curly brace expression. The curly brace expression used was "measurement := READ {openEHR:/query/?aql=…};". Here "aql=" is followed by the AQL expression for getting two blood pressure measurements from a specific patient. The whole curly brace expression with the AQL query can be seen in Figure 3. The MLM then returns the received data. The second MLM contains the logic for orthostatic hypotension (see Figure 1). It was adapted to first call the first MLM for receiving the

needed data, and subsequently check for the conditions of orthostatic hypotension (see Figure 2).

For verifying the "time condition" of the orthostatic hypotension test (within 30 to 180 seconds after standing up) the time of the first measurement was takes as a reference point for standing up.

The connection of the ArdenSuite server to EhrScape was set up using the ArdenSuite FHIR connector, which is capable of sending REST requests and parse the received XML results into Arden Syntax objects. The base URL to the interfaces as well as the login data were entered in order to set up the FHIR connector. After a few minor adjustments, the FHIR connector's internal XML reader could be used to parse the response of EhrScape to form Arden Syntax objects. Therefore, the received data can be processed in our MLMs.

Finally, the MLMs were executed successfully using data received from EhrScape.

4. Discussion

Interoperability between the platforms is useful for both parties: it allows ArdenSuite to use additional data sources for providing its CDS functions, while EhrScape would benefit by the integration of standardized CDS solutions. Further, the interoperability of the Arden Syntax standard and the openEHR standard signifies that both would benefit from a wider field of application.

The "/query" interface of EhrScape was selected for data retrieval because it takes an AQL statement as an input parameter and is therefore the most openEHR-specific one. Additionally, this interface should exist in other implementations of openEHR as well. Steps 3 and 4 of the use case were executed successfully using the test data from Marand's EhrScape. This shows that ArdenSuite is compatible with EhrScape and also indicates its general compatibility with openEHR data.

Using the time of measurement as an indication of when the patient stood up is not an ideal solution for a real operational environment. Instead, a specialized archetype that includes information about the change of position of the patient could serve as a suitable solution.

More openEHR-based systems will have to be tested in order to further ensure the interoperability of ArdenSuite with openEHR. The clinical data repository EtherCis [21] developed by the Ripple Foundation, and the EHRServer [22] from CaboLabs are candidates for further testing because they are open-source openEHR solutions.

The openEHR Foundation is currently developing further specifications for the result structure of AQL queries [23], specifically data fields that should be used to construct the response to an AQL query in any format. This specification should be kept in mind for future adaptations of an openEHR connector for ArdenSuite.

Since the FHIR connector had to be adapted in order to establish the connection to EhrScape, we recommend the development of a specific openEHR connector in addition to the existing FHIR connector, in order to increase interoperability and provide easy-to-use connectors for different purposes.

Curly brace expressions are needed to send AQL requests to the openEHR server, as well as to retrieve and use the returned data. This shows that it is possible to connect these two standards in a way that they can communicate. In the future this may be extended to a more genuine integration of the openEHR approach in the Arden Syntax. This could possibly include the specific use of archetypes or templates in MLMs.

Acknowledgments

The authors are indebted to Katharina Adlassnig, Jeroen de Bruin, Gabriel Kleinoscheg, and Alexander Seitinger for their valuable suggestions, which helped to improve this report. We also thank Borut Fabjan for providing the test access to EhrScape.

References

[1] Medexter Healthcare GmbH, ArdenSuite CDS Platform, https://www.medexter.com/products-and-services/ardensuite, last access: 26 January 2018.
[2] Marand d.o.o., EHRScape, https://www.ehrscape.com/, last access: 13 February 2018.
[3] Health Level Seven International, Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10, November 2014, available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=372, last access: 22 January 2018.
[4] Health Level Seven International, Health Level Seven International – Homepage, http://www.hl7.org/index.cfm, last access: 22 January 2018.
[5] M. Samwald, K. Fehre, J. de Bruin, K.-P. Adlassnig, The Arden Syntax Standard for Clinical Decision Support: Experiences and Directions, Journal of Biomedical Informatics 45 (2012), 711–718.
[6] Health Level Seven International, Arden Syntax, http://www.hl7.org/special/Committees/arden/index.cfm, last access: 22 January 2018.
[7] R.A. Jenders, K.-P. Adlassnig, K. Fehre, P. Haug, Evolution of the Arden Syntax: Key Technical Issues from the Standards Development Organization Perspective, Artificial Intelligence in Medicine (in press); doi: 10.1016/j.artmed.2016.08.001.
[8] openEHR Foundation, Welcome to openEHR, http://www.openehr.org/, last access: 11 January 2018.
[9] openEHR Foundation, What is openEHR?, http://openehr.org/what_is_openehr, last access: 11 January 2018.
[10] K. Atalag, T. Baele, R. Chen, T. Gornik, S. Heard, I. McNicoll, openEHR – A semantically-enabled vendor-independent health computing platform, available at: https://www.openehr.org/resources/white_papers, last access: 15 January 2018.
[11] openEHR Foundation, Archetype Query Language (AQL), : http://openehr.org/releases/QUERY/latest/docs/AQL/AQL.html, last access: 11 January 2018.
[12] openEHR Foundation, OpenEHR REST Definitions API, https://openehr.github.io/specifications-ITS/definitions.html, last access: 23 January 2018.
[13] F. Ricci, R. De Caterina, A. Fedorowski, Orthostatic Hypotension: Epidemiology, Prognosis, and Treatment, Journal of the American College of Cardiology 66 (2015), 848–860.
[14] Medexter Healthcare GmbH, Medexter Healthcare | Home, https://www.medexter.com/, last access: 22 January 2018.
[15] Alfresco Software, Inc., Activiti BPMN Software Home, https://www.activiti.org/, last access: 14 February 2018.
[16] Marand d.o.o., Vendor-neutral, open health data platform | Marand, http://www.marand.com/, last access: 23 January 2018.
[17] Marand d.o.o., API Explorer, https://www.ehrscape.com/api-explorer.html, last access: 26 January 2018.
[18] Marand d.o.o., Login:: ehrExplorer, https://www.ehrscape.com/explorer/, last access: 30 January 2018.
[19] Postdot Technologies, Inc., Postman | API Development Environment, https://www.getpostman.com/, last access: 29 January 2018.
[20] A.E. Aydin, P. Soysal, A.T. Isik, Which is Preferable for Orthostatic Hypotension Diagnosis in Older Adults: Active Standing Test or Head-up Tilt Table Test?, Clinical Interventions in Aging 12 (2017), 207–212.
[21] Ripple Foundation, EtherCIS, http://ethercis.org/, last access: 30.1.2018.
[22] CaboLabs, CaboLabs Health Informatics, Standards and Interoperability, https://www.cabolabs.com/, last access: 30 January 2018.
[23] openEHR Foundation, OpenEHR REST Query API, https://www.openehr.org/releases/ITS/latest/docs/query.html, last access: 29 January 2018.

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

Subject Index

Health Informatics Meets eHealth
G. Schreier and D. Hayn (Eds.)

Author Index